Banking & financial markets

Banking & Financial Markets: Essays in Asset Pricing and Empirical Banking

Michael Kurz

Maastricht University School of Business and Economics

Maastricht

A thesis submitted for the degree of

Doctor of Philosophy at Maastricht University
September 12, 2019

ISBN 978-94-6380-430-1

Cover design by ProefschriftMaken.
Cover picture by Barcin via Getty Images.
Page layout by Michael Kurz.
Printed by ProefschriftMaken.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in Arno Pro, designed by Robert Slimbach in the style of book types from the Aldine Press in Venice, and issued by Adobe in 2007. A template, which can be used to format a PhD thesis with this look and feel, has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/ or from its author at suchow@post.harvard.edu.

BANKING & FINANCIAL MARKETS: ESSAYS IN ASSET PRICING AND EMPIRICAL BANKING

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof.dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday, September 12th, 2019 at 12:00 hours.

by

Michael Kurz

Supervisor

Prof. dr. Jaap Bos

Co-Supervisor

Dr. Paulo Rodrigues

Members of the Degree Committee

Prof. dr. Rob Bauer, Chairman
Prof. dr. Dirk Broeders
Prof. dr. Christian Schlag (Goethe University Frankfurt)
Prof. dr. Nicole Branger (University of Münster)

Contents

Listing of figures

In memory of my grandparents

Edith & Günther Schattschneider

without your support I wouldn’t be where I am today.

Acknowledgments

We are the product of those who taught us, who gave us an opportunity, who have given us chances, who've inspired us.

– Thomas Keller,
Chef

When I read the quote of Thomas Keller above, I could immediately identify with it. Without all the people who I have met in my life so far, those who have taught me and inspired me, I would not be who I am today. Writing this thesis has been challenging but also very rewarding. I could not have completed it without the constant support and guidance of my two supervisors Paulo Rodrigues and Jaap Bos. I learned so much from both of you and I'm grateful for everything that you have both taught me. I'm grateful for the inspiration and many opportunities you have given me. Paulo, you have been not only a PhD supervisor, but also a mentor to me. For the past three years, your door has always been open for me to get feedback on my research ideas, no matter how vague they were, and to get your advice on general issues. I have learned so much from you about econometrics, programming, and about turning a vague research idea into a paper. Jaap, I learned a great deal from you about writing and storytelling and how to presenting my research. You have consistently offered me valuable advice for my future career. Paulo and Jaap, thank you for all of that.

I also feel deeply indebted to Leonard van Effrink. I first came to Maastricht University in August 2014 when I participated in a course in the Maastricht Summer School while still being a master's degree student at Eberhard Karls University of Tübingen. Leonard was the instructor on that course. Leonard took me to lunch with him during that summer. Among other things, we discussed my wish to pursue a PhD after completing my master's degree. Leonard introduced me to Jaap. I had a brief meeting with Jaap one or two days later, and another day later, Jaap introduced me to Paulo. Paulo, Jaap, and I went to lunch and discussed some of my research ideas and my motivation to pursue a PhD. That is how my PhD journey at Maastricht University started.

There are many other people in the Department of Finance at Maastricht University who have taught me and inspired me. I want to thank my two co-authors Stefanie Kleimeier and Rogier Quaedvlieg. I learned a great deal from both of you about econometrics and how to write a good paper. Stefanie, I also enjoyed all the stories you told me about your travels to South Africa and Singapore. I am deeply indebted to Rob Bauer, Dirk Broeders, and Thomas Post, for all the career advice you offered me and all the introductions you made. Stefan Straetmans, I would like to thank you for the many encouraging words over the past three years. I would like to thank Peter Schotman for our discussions about research; every time we spoke, I felt that I had learned something from you. I would also like to thank Dennis Bams, Piet Eichholtz, Rachel Pownell, Jeroen Derwall, and Rogier Otten for our many insightful conversations. I am also grateful to Paul Söderlind for the feedback you gave me on my research during your visit to Maastricht University.

I also would like to thank the members of my PhD degree committee Rob Bauer, Dirk Broeders, Christian Schlag, and Nicole Branger and the additional members of the corona Annelies Renders, Peter Schotman, and Paul Wessels for their time and considerations in assessing this thesis.

During my time at Maastricht University I met many people who became good friends, first and foremost, my two office mates for the better part of my time as a PhD candidate in Maastricht, Aderajew (Addy) Tamirat and Runliang (Jack) Li. I am lucky to have you both as friends, and even though our paths took us to three different continents after our time in Maastricht, I sincerely hope to stay in touch over the years to come. I am also grateful to my paranymphs Ming Li and Joyce Mertens, I feel lucky to call you two my friends. Ming, thank you for all our conversations and for coming to my apartment so many times to ”babysit” my rabbit when I went on vacation. Joyce, thank you for the many conversations over coffee in the staff lounge. Of course, there were many other people with whom I shared my journey as a PhD candidate. Nora, Matteo, Nagihan, Clarissa, Allessandro, Juan, Irene, Lidwin, Mike, Inka, Luuk, Shusen, Hang, Ehsan, Iman, Wiebke, Judy, Martijn, Matthajis, Tobias, Anne, Marina and all the others I may have forgotten to mention; I am grateful that I had the opportunity to meet all of you. Of course, I would also like to thank Francien, Els, and Carina, the Department of Finance’s secretaries. Thank you for the great work that you are doing.

To be honest, I had never felt attached to any place before until becoming part of the Department of Finance because of all the great people I am lucky enough to have met there.

In a way, my PhD journey really started already when I started my master’s degree at Eberhard Karls University of Tübingen. During my two years in Tübingen, I learned more than I had initially imagined I would, and I had the chance to meet so many inspiring people there. I would like to thank my master’s degree thesis supervisor, Rainer Schöbel, for all the advice I received while working on my thesis, but also for the advice regarding my plans for pursuing a PhD outside of Germany. My time in Tübingen would also not have been the same without the friends I made along the way, specifically Konstantin, Uli, Mandy, and Amber.

Of course, most importantly, I met my wife Nhung while we were both studying in Tübingen. Nhung, you have always been there for me, you always stand by my side to support me and give invaluable advice or simply listen to me. You inspire me more than any other person in my life. I am happy that our paths crossed in Tübingen. Thank you for everything.

Lastly, I would like to thank my family, especially my mother, my brother Marcel, and my grandparents Edith and Günther Schattschneider, to whom I dedicate this thesis. Without your support and advice, I would not be who I am today.

I have no doubt that I inadvertently left people out who deserve to be thanked. Whether I mentioned you or not, be sure that our memories won't be forgotten, and your friendship is truly cherished.

Michael Kurz
Almere, September 12th, 2019

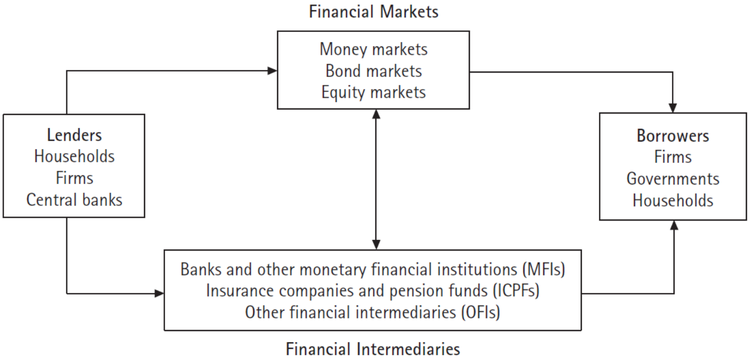
1 Introduction

BANKS AND FINANCIAL MARKETS are interconnected in many important ways. The 2007-2009 financial crisis serves as a reminder of the important role of banks, financial markets, and the interconnection between them in the global economy. Historically, banks and financial markets have been considered competitors in the financial system (see, e.g., Allen and Gale (1997), Boot and Thakor (1997)). Recently, Song and Thakor (2010) and Bossone (2010) have suggested a view of the interconnection between banks and financial markets, according to which both realms co-evolve and compete, but also complement each other.

Figure 1.0.1 presents a stylized overview of the financial system and the connections between banks and financial markets. Those in the economy who possess excess funds (Lenders) provide funding to those in the economy who need financing (Borrowers) either

Figure 1.0.1: Banks & Financial Markets in the Financial System

Source: Allen, Chui, and Addaloni (2004), page 491.



through banks or through financial markets. However, banks themselves also depend on financial markets as a source of their own funding, for hedging purposes or to engage in proprietary trading, market-making, and securitization (see, e.g., Allen and Carletti (2009)). The chapters of this thesis are mostly concerned with the direct connection between banks and financial markets that is indicated by the arrows in the center of Figure 1.0.1. The dimensions of banks direct interact with financial markets may serve to reduce risk (in case that a bank engages in hedging) or to elevate risk (in case that a bank engages in proprietary trading) (see, e.g., Boot and Thakor (2009)). In Chapters 2 and 3, these aspects of interconnection between banks (dealing in securities for the purpose of either hedging or proprietary trading) and financial market interconnection are investigated more closely from different perspectives.

Credit Supply and Proprietary Trading

Chapter 2 explores the historically controversial relationship between the proprietary trading business and the lending business of universal banks. Universal banks combine classical commercial banking services, such as lending and payment services, with a wider range of financial services, such as securities underwriting and trading (see Morrison (2009)). The question of whether or not banks that engage in classical commercial banking activities should also be permitted to engage in proprietary trading has been discussed by financial economists, legal scholars, and policymakers for a long time. For example, since the late nineteenth century, the US has repeatedly switched between a system of universal banking and one that separates commercial banking from securities underwriting and trading.[1]

While relying on implicit or explicit government guarantees for cheap funding, banks may find it more profitable to invest in trading operations rather than their lending business. In particular, during a crisis, banks could be inclined to purchase securities for fire-sale prices speculating on future returns during economic recovery rather than providing loans to non-financial firms. On the one hand, such behavior by banks can support financial markets through the provision of liquidity. On the other hand, it could lead to a spillover of security price shocks to the real economy in the form of a credit crunch.[2].

Chapter 2 primarily addresses the question: "Do banks that heavily engage in proprietary trading reduce credit supply in times of crisis relative to other banks?" The results reported in Chapter 2 suggest that banks with greater trading expertise indeed supply less credit during stable times and even less during times of crisis. Compared to non-trading banks, trading banks reduce their credit supply by 19% plus an additional 3.25% during crises. This double effect can be attributed to US banks. International banks are unique in this regard as they only reduce their credit supply during crises. These spillovers from trading to credit supply have adverse consequences for the real economy as firms are weakened in their ability to invest in capital and expand their workforce.

Financial Derivatives and Accounting Rules

Banks may turn to financial markets to buy or sell financial derivatives to use them either for hedging or proprietary trading. Randall Dodd called derivatives a "double-edged sword" since "they are extremely useful for risk management, but they also create a host of new risks that expose the entire economy to potential financial market disruptions"(see Berry (2003)). Indeed, banks have repeatedly incurred substantial losses due to their dealings in derivatives. For example, A.I.G. suffered $18 billion in derivatives-related losses

[1]See D'Arista (1994) for an excellent historical overview of the development of the US banking system, including comparisons with the historical development of the banking systems of other countries.

[2]See Shleifer and Vishny (2010), Diamond and Rajan (2011), Arping (2013), Stein (2013), Boot and Ratnovski (2016)

in 2008, and Morgan Stanley and Société Générale lost $9 billion and $7.2 billion respectively in the derivatives market in the same year. Earlier examples of financial institutions that incurred significant losses due to their use of derivatives include Allfirst Bank ($691 million in 2003), Daiwa Bank ($1 billion in 1997), Barings Bank ($1.4 billion in 1995), or Midland Bank ($500 million in 1993). These instances exemplify the need of investors and regulators to properly understand how banks use derivatives. Nevertheless, the reporting of derivatives use is exceedingly complex, leading to intransparency (see Leone (2007), Valladares (2014), Chang, Donohoe, and Sougiannis (2016)). Survey evidence suggests that reporting incentivizes banks to under-report the extent of their hedging activities (see Mulford and Comiskey (2008), Papa and Peters (2013)).

However, Bushman (2016) defines bank transparency as the "availability to outside stakeholders of relevant and reliable information about periodic performance, financial positions, business model, governance, and risks of banks". Chapter 3 relies on a latent class regression model to estimate the probabilities of banks using derivatives primarily for hedging or for trading, independent of their reported derivatives use. Consequently, banks are classified as hedgers or traders based on the highest probability of class membership.[3] To avoid reliance on reported derivatives use, this approach is based on the notion that, with everything else equal, a bank that engages in hedging should have, on average, a lower probability of default than a bank that engages in trading. In this way, the chapter aims to increase the transparency of banks' derivatives use.

Our results also show that while there is an overall tendency of banks to use derivatives for hedging purposes, those banks' with derivative holdings far above the sample average tend to use derivatives for speculative trading. In particular, during the 2007-2009 financial crisis, the average class-membership probabilities of the hedger class declined, but the proportion of derivatives held by banks increased. During the peak of the crisis, we observe the largest number of US banks changing from hedging to trading or from trading to hedging. The banks that changed from hedging to trading during the crisis experienced a

[3] The probability of class membership is the probability of "observing" the current z-score (i.e. probability of default) conditional on the bank being a hedger or trader. For example, Bank A may have a z-score of 2 in the current quarter. Based on a large number of bank characteristics the latent class regression model may estimate that the probability of Bank A having a z-score of 2 in the current quarter is 5% if Bank A uses derivatives primarily for trading and 95% if it uses derivatives primarily for hedging. In this case Bank A would be classified as hedger.

sharp decline in their return-on-assets prior to the change, consistent with the banks' attempts to increase risk in order to boost profitability. The banks that changed from trading to hedging during the same time period exhibited rather stable return-on-assets despite the financial crisis, consistent with such banks wanting to "lock-in" current income.

Estimating Betas and Risk Premiums

Besides dealing in securities, banks may also turn to financial markets to raise capital to fund their own operations. Therefore, banks' funding costs are affected by market frictions and behavioral anomalies documented in the empirical asset pricing literature. Chapter 4 of this thesis is not about banking per se but discusses issues in the OLS estimation of CAPM betas and associated market risk premiums due to the influence of various statistical biases. These statistical biases in the estimation of CAPM betas and risk premiums have direct implications for the discussion concerning the appropriate leverage levels of banks.

The 2007-2009 financial crisis revealed the substantial economic costs of highly leveraged banks. In response, governments in almost all major economies set new capital requirements for banks, forcing them to use more loss-absorbing equity capital. Since equity has a higher required return than debt, this has led to concerns that forcing banks to increase equity would increase banks' funding costs.[4] In competitive lending markets, this in turn is reflected in the higher interest rates charged to those who borrow from banks. However, from a theoretical perspective, there are compelling arguments to refute the claim that more equity implies higher funding costs.

In efficient financial markets, asset prices reflect all publicly available information. Therefore, there is always an elastic supply of capital available to banks at a price that reflects their fundamental values (see, e.g., Baker (2009)). While for shareholders of banks with higher equity the return-on-equity is lower during a good economic state (when return-on-assets is high), the return-on-equity is higher during a bad economic state (when return-on-assets is low).[5] Thus, if a bank uses relatively more equity, its resilience to a bad economic state increases, and it therefore provides a downside protection that reduces

[4] See, e.g., Admati, DeMarzo, Hellwig, and Pfleiderer (2011)

[5] To understand why, recall that the return-on-equity (ROE) can be written as $ROE = ROA + (D/E)(ROA - r)$, where $ROA = EBIT(1 - Tax\ Rate)/A$, A denotes total assets, E and D are equity and debt respectively, and r is the after-tax interest rate paid on D (see Admati et al. (2011)).

shareholders' risk. With the higher equity share, the banks' sensitivity to the risk of assets – and therefore the banks' "beta" – declines, leading investors to demand a lower risk premium and hence a lower required average return for holding banks' equity. Because more equity also translates into a lower probability of default, cost of debt for banks should decline. Therefore, basic corporate finance theory tells us that if financial markets are efficient, increasing equity should lead to a lower cost of debt and a lower required average return for holding banks' equity, as more equity implies lower risk. Even if we acknowledge that equity funding is costlier than debt funding, the weighted average cost of the banks' capital may be lower for banks using more equity (see, e.g., Miles, Yang, and Marcheggiano (2012), Admati, DeMarzo, Hellwig, and Pfleiderer (2011)).

However, it is conventional wisdom that financial markets are not efficient; rather, they are plagued with the varied frictions and behavioral biases of market participants. Of particular interest to the preceding discussion on the effect of increased equity on funding costs is the "low risk anomaly" – sometimes also called the "beta anomaly". This anomaly is derived from the empirical observation that in stock return data, it appears that risk, as measured by the CAPM beta, is negatively related to average returns (see, e.g., Frazzini and Pedersen (2014), Ang (2014), Baker, Bradley, and Wurgler (2011)). In the presence of this anomaly, the previous discussion takes an about-face: As banks use more equity, they become less risky, but because of their lower risk, investors demand higher rather than lower average returns for holding the banks' equity. Baker and Wurgler (2015) demonstrate that if there is a low risk anomaly, then one percentage point increase in a bank's equity share would lead, on average, to an increase in the weighted average cost of capital of 85 basis points per annum. This could result in a substantial increase in the interest rates charged to the banks' clients.

Therefore, Chapter 4 is concerned with the question of whether the relationship between the CAPM beta and average returns is negative or positive. If this relationship is positive, investors demand (and receive) a positive premium for holding market risk, as measured by the beta. Furthermore, banks should not face the increasing weighted average cost of capital when they use more equity, since more equity should reduce the banks' beta. However, if we want to know whether there is a "low risk anomaly" for banks, it may not be sufficient to only analyze bank stocks. Therefore, in Chapter 4, a broader view is taken: In fact, there is a positive relationship between the beta and average returns.[6] Thus, while Chapter 4 does not discuss banking per se, the conclusions thereof have direct implications for the assessment of banking regulation.

The results in Chapter 4 demonstrate that betas are difficult to estimate precisely, since such estimates are affected by three sources of statistical biases: price measurement error, sampling error, and time-series variation in betas. If betas are estimated using OLS, the measurement error in the underlying return data due to price staleness and illiquidity, the sampling bias due to the small sample for estimation, and the time-series variation in true betas lead to noisy beta estimates. These error are not independent from each other and, therefore, they produce non-trivial trade-offs between bias and variance. These statistical biases (rather, frictions and behavioral biases) lead to the appearance of a "low risk anomaly" in stock markets. Therefore, our empirical results let us conclude that the negative relationship between the CAPM beta and average returns is a measurement issue rather than an anomaly.

[6] Originally, Chapter 4 began as a research project entirely focused on the cost of capital of banks in the presence of asset pricing anomalies in general and the low risk anomaly in particular. I would like to thank Luis Viceira for suggesting a switch of focus in the chapter from banking towards the more fundamental questions that the chapter now tries to answer.

INTRODUCTION

2

Credit Supply: Are there negative spillovers from banks' proprietary trading?[1]

2.1 Introduction

Traditionally, banks are described as institutions that primarily accept deposits from households and provide loans.[2] However, the business model of most large modern banks extends beyond commercial banking, as banks are heavily involved in financial markets

[1]This chapter is based on a working paper co-authored by Stefanie Kleimeier (Maastricht University; Open Universteit Heerlen; Stellenbosh Business School)

[2]For example, the IMF writes in its Finance & Development series "Back to Basics" that "[Banks'] primary role is to take in funds - called deposits - from those with money, pool them, and lend them to those who need funds". See Gobat (2012).

through the origination, distribution, and trading of various kinds of securities. Since the 2007 financial crisis, the proprietary trading activities of banks have come under great scrutiny. The Volcker Rule in the US, the Vickers Report in the UK, and the Liikanen and European Commission proposals in the EU all aim to limit the risks believed to emanate from banks' trading activities by strictly separating trading from commercial banking business.[3] The concern underpinning these rules is that banks take on large risky bets while relying on implicit or explicit government guarantees for cheap funding, and then threaten to discontinue to offer classic banking services. In particular, during a crisis, banks could be inclined to purchase securities for fire-sale prices, speculating on future returns during economic recovery, rather than providing loans to non-financial firms. On the one hand, such behavior by banks can support financial markets through the provision of liquidity. On the other hand, it could lead to a spillover of security price shocks to the real economy in the form of a credit crunch.[4] Based on these considerations, we test two hypotheses: First, we test the hypothesis that banks with greater trading expertise supply the real economy with less credit relative to banks with lower trading expertise, especially during periods of crisis. Second, we test whether this lower credit supply leads to lower investments and lower employment growth in non-financial firms that depend on funding from banks with trading expertise. We use a global sample of bank-firm lending relationships along with firm- and bank-specific information covering 135 banks from 21 countries and their lending to 8,242 firms from 81 emerging and advanced economies over the period 2003 to 2016. We find evidence in support of both hypotheses, suggesting that regulators' concerns regarding proprietary trading are generally well founded. Hence, the regulations are an important and justified tool of economic policy, despite some negative implications for market-making and liquidity. We provide evidence for the existence of the negative real economic effects of proprietary trading that need to be taken into account by regulators when assessing the cost-benefit trade-off of the above regulations.

We contribute to the literature on this topic by analyzing a global sample of corporate loans, from 2003 to 2016, using the Thomson Reuters LPC DealScan database which we

[3]See Lehmann (2016), Krahnen, Noth, and Schüwer (2017)

[4]See Shleifer and Vishny (2010), Diamond and Rajan (2011), Arping (2013), Stein (2013), Boot and Ratnovski (2016). Besides spillovers, there are also concerns regarding the conflict of interest of banks engaging in proprietary trading and simultaneously advising clients on trading. For a comprehensive discussion on the US context, see, e.g., Merkley and Levin (2011).

hand-match with Standard & Poor's Compustat database to obtain bank and borrower characteristics. This allows us to also provide an estimate of the real economic effects in terms of investments and employment growth. More specifically, we show that banks that are heavily engaged in securities trading supply non-financial firms with roughly 19% less credit compared to banks less heavily engaged in securities trading. That gap in credit supply increases even further during periods of crisis. We further demonstrate that the reduced credit supply of banks heavily engaged in securities trading has ramifications for their borrowers. Firms tend to invest less in capital and expand their workforce at a lower rate the more they depend on trading banks for financing. Moreover, our results indicate that while trading banks generally charge their borrowers higher spreads, they do not increase loan prices beyond what is observed from their non-trading peers during a crisis. By examining our global sample, we also find that while trading banks provide less credit overall, they tend to provide slightly more credit abroad. However, during a crisis, trading banks also cut their foreign lending to a greater extent than their non-trading peers. Finally, we show that there are significant differences between US banks and other banks in our sample. Non-US banks that are heavily engaged in securities trading only reduce their credit supply during a crisis, but not during economically stable times. However, for US banks that are heavily engaged in securities trading, we find a reduction in credit supply both during times of crisis and stability.

Our empirical analysis tests predictions from a large base of theoretical literature on the role of banks' securities trading. Shleifer and Vishny (2010) and Diamond and Rajan (2011) argue that if funds are scarce, banks with greater trading expertise may reduce credit supply during a crisis as they redirect funds from lending to trading as the returns from investing in distressed assets are higher than the returns from lending. Arping (2013) makes a similar point and shows that while this behavior is individually optimal for banks from a profit-maximization perspective, it may hamper growth in the real economy as non-financial firms find it increasingly difficult to obtain credit financing. Even beyond periods of crisis, Boot and Ratnovski (2016) show in a theoretical model that the allocation of scarce funds to scalable short-term securities trading tends to reduce the availability of credit for non-scalable long-term relationship lending activities. This reallocation leads to insufficient incentives for banks to build and maintain long-term lending relationships. Moreover, Krahnen, Noth, and Schüwer (2017) point out that universal banks that calculate their funding costs by averaging over the (high-risk) funding costs of trading and the (low-risk) funding costs of lending rather than using separate funding costs are biased towards trading. Averaging funding costs leads to a relative change in the profitability of trading and lending activities, where trading profitability increases as the average funding costs are below the trading funding costs and lending profitability decreases in an off-setting manner, as the average funding costs are higher than the lending funding costs. Hence, banks would be incentivized to cut funds for lending while increasing funds for trading.[5] Neither of the described theoretical models nor the above-cost averaging argument depend on government guarantees for bank liabilities. However, in the presence of government guarantees, additional incentives would be provided for banks to increase their trading activities at the expense of relationship banking, as the funding costs of trading activities will not fully reflect the investment risks.[6] In summary, all these theories support our first hypothesis, while Arping (2013) supports our second hypothesis.

A large base of empirical literature documents the declining credit supply during the 2007-2009 financial crisis (see, e.g., Ivashina and Scharfstein (2010)). This decline is traced back to different bank lending channels, among which the most prominent is the bank lending channel of monetary policy (see, e.g., Bonaccorsi di Patti and Sette (2012), Jimenez, Ongena, Peydro, and Saurina (2012), Jimenez, Ongena, Peydro, and Saurina (2014)). Additionally, Cingano, Manaresi, and Sette (2016), and Iyer, Peydro, da Rocha-Lopes, and Schoar (2014) consider shocks to interbank lending, while Acharya, Eisert, Eufinger, and Hirsch (2018) investigate the role of banks' exposure to crisis-induced sovereign risk thorough bond holdings. Finally, Abbassi, Iyer, Peydro, and Tous (2016) investigate the trading channel that also drives our analysis. For the German banking market, Abbassi et al. (2016) show that those banks with trading expertise reduce lending more than banks without trading expertise during the financial crisis and redirect internal funds to buy stressed assets for fire-sale prices. The authors have access to a unique central bank dataset, including the German credit register and the European security-level holdings database, and are therefore able to provide security-specific evidence on trading decisions.

[5]Krahnen et al. (2017) argue that this could easily be avoided through appropriate internal transfer pricing, but bank managers have personal incentives to apply cost averaging in the described way. This means that if managers have access to a bonus pool from one segment (either trading or banking) but not from another segment, cost averaging may increase the income of the managers at the expense of the shareholders.

[6]See Krahnen et al. (2017).

While Abbassi et al. (2016) focus mostly on trading, they also provide some evidence that higher trading expertise goes hand-in-hand with lower credit supply provided by German banks to German firms. Thus, the global dimension of the trading-credit supply link and its real economic effects remain unexplored. Our analysis fills this gap.

2.2 Empirical Framework

Our aim is to investigate whether banks with extensive trading operations provide fewer loans in the corporate loan market than banks without or less extensive trading operations. To answer this question, we apply a modified version of the Khwaja and Mian (2008) regression specification. Khwaja and Mian (2008) consider an economy in which firms borrow from multiple banks. Such an economy may experience two kinds of observationally equivalent shocks to bank lending: firm-specific credit demand shocks and bank-specific credit supply shocks. Credit demand shocks reflect unobserved changes to firms' fundamentals such as shocks to productivity or shocks to customer demand. Credit supply shocks reflect changes in banks' funding situation such as variations in the availability of deposits or short-term liabilities or, as is the focus in this chapter, redirection of available funds from corporate lending to proprietary trading. Therefore, it necessary to use an econometric specification that allows us to isolate the relevant credit supply shock.

The main idea of the Khwaja and Mian (2008) approach is the use of matched bank-borrower data to achieve this by controlling for unobserved credit effects to identify supply effects. Initially, we estimate the following model:

$$\begin{aligned}\Delta Log\left(LoanVolume\right)_{ijt} = \beta \times Trading_i + \varphi \times FSI_{it} + \xi \times \left(Trading_i \times FSI_{it}\right) \\ + \boldsymbol{\delta} \times \mathbf{X_{it}} + \gamma_{jt} + \gamma_{bank\ country\ t} + \nu_{ijt}, \qquad (2.1)\end{aligned}$$

where the dependent variable is the change in the logarithm of the loan volume by bank *i* to borrower *j* in year *t*. While Equation (2.1) is represented in reduced form, Khwaja and Mian (2008) show that it can be derived as an equilibrium condition by explicitly modeling credit supply and demand schedules. We include borrower*time fixed effects γ_{jt} to account for time-varying, unobserved heterogeneity in borrower characteristics that proxy for credit demand. Hence, we compare the changes in the loan volume extended to the same borrower in the same year across different banks. Our specification also includes bank country*time fixed effects $\gamma_{bank\ country\ t}$ to account for time-varying macroeconomic conditions and regulatory environments in the banks' country of incorporation. Moreover, we include a vector $\mathbf{X_{it}}$ of bank control variables in our model where $\boldsymbol{\delta}$ denotes the corresponding vector of the regression coefficients.

Our coefficient of interest is β, where $Trading_i$ proxies for bank i's trading expertise. In line with our first hypothesis, we expect β to be negative, indicating that banks with greater trading expertise reduce credit supply to their corporate borrowers. We expect this effect to be stronger during periods of crisis than in periods of stability. We therefore interact $Trading_i$ with a Financial Stress Indicator FSI_{it} and expect a negative value for ξ.

Besides the financial effects, i.e., the effects of banks' trading activities on the loan volume granted, we also investigate the real economic effects of banks' trading activities. Following the approach in Khwaja and Mian (2008), Acharya et al. (2018), and Cingano et al. (2016), we estimate the following model:

$$\begin{aligned} y_{jt} = \varphi_{Country} + \varphi_{Industry} + \varphi_t + \theta \times \overline{\text{Exposure}}_{jt} + \Phi \times FSI_{jt} \\ + \lambda \times \overline{\text{Exposure}}_{jt} \times FSI_{jt} + \boldsymbol{\Psi} \times \mathbf{F_{jt-1}} + u_{jt}, \qquad (2.2) \end{aligned}$$

where y_{jt} refers either to the capital expenditure or employment growth of borrower j in year t. $\mathbf{F_{jt-1}}$ is a matrix of the borrower control variables. $\varphi_{Country}$, $\varphi_{Industry}$, and φ_t denote the country, industry, and year dummies respectively. $\overline{Exposure}_{jt}$ is a proxy for exposure of a borrower to the trading expertise of its lender banks. In line with our second hypothesis, we expect θ to be negative, indicating that firms with a greater dependency on trading banks suffer from a more restrictive credit supply and thus exhibit lower capital expenditures and employment growth.

There are different channels through which exposure to trading by lenders can affect borrowers, and we differentiate between the three channels in our empirical model:

$$\overline{\text{Trading Exposure}}_{jt} = \sum_i \omega_{jit} \times 1_{[TradingExpertise_i]} \tag{2.3}$$

$$\overline{\text{Trading Bank FSI Exposure}}_{jt} = \sum_i \omega_{jit} \times 1_{[TradingExpertise_i]} \times FSI_{it} \tag{2.4}$$

$$\overline{\text{Bank FSI Exposure}}_{jt} = \sum_i \omega_{jit} \times FSI_{it} \tag{2.5}$$

where ω_{jit} is equal to the share of credit granted by each bank i to borrower j in year t and $1_{[TradingExpertise_i]}$ equals one if bank i is considered as having trading expertise and zero otherwise. FSI_{it} is a Financial Stress Index, measuring the level of stress in the financial market of bank i's country of incorporation. Thus, $\overline{Trading\ Exposure}_{jt}$ is simply the share of loans granted to a borrower by banks with trading expertise. Meanwhile, by using $\overline{Trading\ Bank\ FSI\ Exposure}_{jt}$ we can capture exposure to financial market stress in the country of incorporation of the lender banks with trading expertise. Finally, with $\overline{Bank\ FSI\ Exposure}_{jt}$, we have a measure of exposure to financial stress in a bank's country of incorporation, unconditional on trading expertise.

Each of the measures in Equations (2.3) to (2.5) captures a different channel by which non-financial firms could be affected by the capital market operations of their banks. Equation (2.3) captures the direct effect of exposure to banks with trading expertise. Equation (2.4) uses the same exposure but further weights it by the current condition of the bank's home financial market. This equation clearly captures the idea that banks with trading expertise would buy assets at fire-sale prices in times of financial market stress.[7] The last measure in Equation (2.5) moves away from the idea of explicitly discriminating between banks with or without trading expertise simply by capturing exposure to the current condition of the bank's home financial market.

2.3 Data

To estimate the two models, we need information on the banks' lending and trading activities as well as the borrowers' exposure to their lenders' trading activities. Our primary

[7] See Abbassi et al. (2016), Diamond and Rajan (2011), Arping (2013).

sources of information are the Thomson Reuters LPC DealScan database, which provides extensive coverage of the global corporate loan market, and Standard & Poor's Compustat database, which provides extensive information on bank and borrower characteristics. Since LPC DealScan and Compustat do not share any common identifier, we hand-match all borrower, bank, and loan information. We collect information on corporate loans extended by 136 major banks in 21 countries between 2003 and 2016 to 8,242 non-financial firms in 81 countries, including advanced and emerging economies.[8] Our banks are based in the US, Canada, the UK, Austria, Belgium, Denmark, France, Germany, Ireland, Italy, Netherlands, Spain, Sweden, Switzerland, China, Hong Kong, Taiwan, South Korea, Singapore, Japan, Brazil, and Australia.

In model (2.1), our dependent variable is the change in the logarithm of the loan volume by bank i to borrower j in year t. In most uses of the Khwaja and Mian (2008) specification, the left-hand side variable is measured using detailed information from national credit registers. However, our corporate loan data differs from such credit register data in two important ways. First, we cannot observe changes in a particular loan over time, since we only observe loans at the time of their origination. Second, the loans in our sample tend to have long maturities. These two properties of our loans imply that for a large number of firms, there is no meaningful time-series variation in the bank-firm loan volumes. To address this issue, we follow Acharya et al. (2018) and aggregate firms into clusters, applying the Khwaja and Mian (2008) estimator to a panel of bank-firm cluster relationships.[9,10]

Similar to Acharya et al. (2018), we form clusters based on the country of incorporation, the two-digit SIC code, and the median EBITDA interest coverage ratio. We expect firms

[8]Consistent with the literature, we aggregate all loans to each bank's parent company (see, e.g., Sufi (2007)) and track bank mergers over our sample period (see, e.g., Schwert (2018)).

[9]Veredas and Petkovic (2010) have shown that aggregating individual observation into groups into panel datasets with a low-time frequency does not affect the model structure. The estimated coefficients remain unbiased and correspond to the coefficients of the individual firm-level regressions. However, heteroscedasticity is introduced due to the aggregation of individual firms. Both statements are easy to verify using standard arguments (see 2.A.2). Thus, for model (2.1), we cluster standard errors at the bank and firm-cluster level, and for model (2.2), we cluster standard errors at the firm-cluster level.

[10]Aggregating individual observations into groups may also raise concerns regarding the Simpson's paradoxon (see Simpson (1951), Blyth (1972)), i.e., the phenomenon that a trend may appear within groups of the data but reverses if the individual observations in the groups are aggregated. However, the inclusion of group fixed effects that act as group-specific intercepts in our regression models prevents trends in the groups from reversing after aggregation of the observation.

that are incorporated in the same country and that are active in the same industry to share sufficiently similar characteristics. Furthermore, credit ratings are an important determinant in bank lending. Therefore, firms with the same rating will have similar access to the loan market or other sources of financing (see, e.g., Diamond (1991), Erel, Julio, Kim, and Weisbach (2011)). Thus, we further match firms in the country-industry clusters based on their median interest coverage ratio. In summary, our dependent variable $\Delta log(LoanVolume_{ijt})$ is the change in the logarithm of the total USD volume of credit granted by bank i to all firms in cluster j in year t. This leaves us with 24,056 unique bank-firm cluster connections.

Our main independent variable is $Trading_i$, which reflects the trading expertise of bank i. Consistent with the approach used in Abbassi et al. (2016), we rely on the notion that banks, in order to maintain or build a strong presence in securities trading and thus to have trading expertise, require a specific infrastructure. Arguably, direct trading memberships at important securities exchanges are among the most relevant aspects of such trading infrastructure, as they allow for direct access to the trading floors and trading and clearing systems of the respective exchanges without the need of intermediate brokers.

Thus, for each bank in our sample, we count the total number of trading memberships at Euronext (the European multi-country exchange), the London Exchange, NYSE, NASDAQ, the Toronto Exchange, the Japan Exchange (covering all Japanese exchanges), the Hong Kong Exchange, the Shanghai Exchange, BMnF Bovespa (Brazil), the Australian Securities Exchange, and the Deutsche Börse (the German Exchange). Each of these exchanges has been listed as one of the ten largest exchanges in terms of market capitalization at least once during our sample period. A bank is considered a trading member of either of these exchanges if it has purchased the right to directly access the trading floor. If a bank has access to more than one market of the same exchange (equity, fixed-income, and/or derivatives), we count this as one membership at the relevant exchange. Note that it is not necessary for foreign banks to possess a banking license in the relevant country to purchase a membership.

We hand-collect the trading membership information from the websites of the relevant exchanges and company reports. While all banks in our sample offer trading services to their clients, it is not necessary for a bank to possess a trading membership at any exchange to offer such services. Such a bank could handle all trading, including trading on behalf of clients via external broker-dealers. Even if a bank were to purchase a trading membership to more easily offer trading services to clients, this would hardly require more than a single membership at a major exchange. Thus, we would expect banks with a large number of exchange memberships to have strong trading operations and consequently a higher level of trading expertise. Therefore, next to a simple count variable of the number of exchange memberships as a proxy for trading, we consider two dummies: one identifying banks with at least one membership, the other identifying banks with more than two memberships. The latter category reflects the idea that banks with only one or two memberships use those primarily for client-related trading, while true proprietary traders require a larger number of trading memberships in various markets.

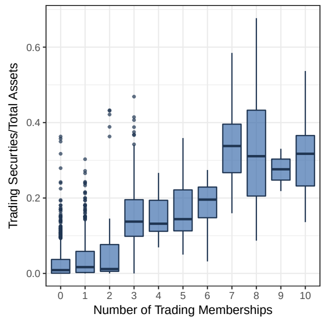
Consistent with our line of argument, Figure 2.3.1 additionally indicates that the USD volume of a bank's trading account as a fraction of its total assets tends to be larger the more trading memberships the bank possesses. We estimate a correlation coefficient of 0.6 between the two variables, which is statistically significant at the one percent level. A larger trading account volume indicates greater securities trading on part of the the banks.

Our notion of interpreting greater activity as a sign of greater expertise in trading is consistent with a large base of theoretical and empirical literature on organizational learning-by-doing (see, e.g., Jarmin (1994), Thompson (2010), Argote and Miron-Spektor (2011)). Note that for all panels of Figure 2.3.1, there is an upward jump in trading securities for banks with more than two memberships. This supports the previously outlined approach of defining a trading expertise dummy that equals one if a bank has more than two memberships and zero otherwise.

Ideally, we would like to observe when banks are buying or selling securities. Changes in the USD volume of a bank's trading account cannot be used to identify when banks are trading. Since the volume is the product of price and quantity, increases in quantity due to banks' purchases of securities could be offset by the prices of the same securities falling during crisis.

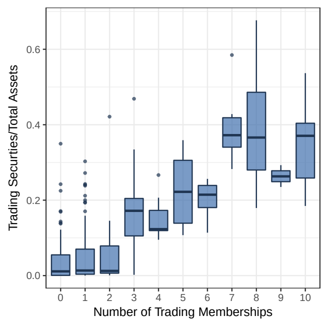
Figure 2.3.1: Trading Account and Trading Memberships

Notes: In the boxplot, we show the volume of the securities trading account as a fraction of the total assets for different counts of trading memberships at exchanges. The sample consists of 136 major banks based in 21 countries between 2003 and 2016. Panel a) shows the boxplot for the full sample period, and panels b) to d) show the boxplots for the various sub-periods. The continuous variable (y-axis) represents the USD volume of the trading/dealing account divided by the USD (book value) of the total assets. The categorical variable (x-axis) represents the number of trading memberships at major exchanges. We count memberships at Euronext (the European multi-country exchange), the London Exchange, NYSE, NASDAQ, the Toronto Exchange, the Japan Exchange (covering all Japanese exchanges), the Hong Kong Exchange, the Shanghai Exchange, BMnF Bovespa (Brazil), the Australian Securities Exchange, and the Deutsche Börse (the German Exchange).

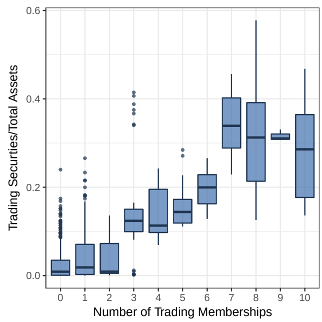


a) Full sample period

b) Pre-Crisis period



c) Crisis period



d) Post-crisis period

With respect to FSI_{it}, we consider three different measures. First, we consider a simple dummy variable – *Crisis* – indicating the crisis period from 2007 to 2009. Note that if we include the crisis dummy, we must adapt the specification of the fixed effects in model (2.1). Accordingly, we use firm cluster and bank country fixed effects rather than firm cluster-year and bank country-year fixed effects. Second, we augment our dataset with the Financial Stress Index – $\textit{Financial Stress}_{it}$ – developed by the US Office for Financial Research.[11] The Financial Stress Index is a continuous measure of stress in financial markets, taking into account contributions to stress from credit markets, equity valuations, funding, safe assets, and volatility. The index is centered on zero, where positive values indicate increased stress and negative values indicate relaxation. Thus, using the index, we can obtain a more granular view of the financial market conditions over time compared to a simple crisis dummy. Furthermore, since the index distinguishes three different world regions (the US, other advanced economies, and emerging economies), we can take into account that emerging economies were less affected by the 2007-2009 financial crisis than advanced economies were. This impact is documented in, e.g., Blanchard, Das, and Faruqee (2010). In particular, Emerging Asia was affected to a lesser extent than advanced economies were (see, e.g., Goldstein and Xie (2009), Keat (2009)). Thus, we feel it is important to take these differences into account.

We use price and market capitalization data from Standard & Poor's Global Equity country indices to construct a measure of capital market conditions:

$$\begin{aligned} \textit{Capital Market}_{it} = & \\ & -1 \times \left(\frac{MCap_{Home,t} \times Price_{Home\ t} + \sum_{k=1,k\neq Home}^{K} MCap_{kt} \times Price_{kt}}{MCap_{Home\ t} + \sum_{k=1,k\neq Home}^{K} MCap_{kt}} \right) \end{aligned} \tag{2.6}$$

where $MCap_{kt}$ is the total market capitalization of country k's stock market, and $Price_{kt}$ is the value of the Standard & Poor's Global Equity Index for country k. The k indexes countries in which bank i possesses trading memberships and *Home* indexes a bank's country of incorporation. We multiply the right-hand side of the equation by -1 to obtain the same directionality as the previously described Financial Stress Index, i.e., due to the multiplication by -1, high values of $\textit{Capital Market}_{it}$ indicate low prices and vice versa.

Finally, the bank control variables $\mathbf{X_{it}}$ in model (2.1) capture differences in bank size,

[11] See Monin (2017) for details on the computation of the index. The US Office for Financial Research was created by the Dodd-Frank Act of 2010 and is tasked with observing US and global financial markets conditions to provide regulators with timely market intelligence.

profitability, and funding. The control variables comprise the logarithm of the book value total assets, ROA, capital ratio, liquidity ratio, and the Loans-to-Deposits Ratio. The data for these controls is obtained from Compustat. ROA is computed as the income before extraordinary items divided by the book value of total assets. The capital ratio is the ratio of the book value of common equity to the book value of total assets. The liquidity ratio is computed as the ratio of cash to total assets. The Loans-to-Deposits Ratio is computed as the ratio of total loans to total deposits.

In model (2.2), our dependent variables y_{jt} are the capital expenditure and employment growth in year t aggregated across all firms in cluster j. Our firm cluster controls $\mathbf{F_{jt-1}}$ comprise the logarithm of the book value of total assets, net debt-to-assets ratio, intangible assets-to-assets ratio, the change in cash and cash equivalents in year $t-1$ aggregated across all firms in cluster j. Net debt is the sum of short-term and long-term liabilities minus cash and cash equivalents. The inputs for our three exposure measures are derived from the trading expertise measures of model (2.1). Looking ahead to our results for model (2.1), we find that our simple dummy that indicates more than two trading memberships is most informative. We therefore utilize this dummy in our implementation of model (2.2).

2.4 Summary Statistics

We show summary statistics for our sample banks in Table 2.4.1. The banks in our sample are large and rather similar in size. However, there is significant variation in both profitability measured by ROA and capitalization. The large variations in Total Loans and Trading Securities suggest differences in the business models of our sample banks. For the average bank in our sample, loans account for roughly 50% of assets and trading securities account for roughly 9% of assets. However, there are banks with particularly large holdings in trading securities.

The number of trading memberships varies from non membership to memberships at all of the exchanges considered in our analysis, and the average bank possesses two memberships. The bottom row in Table 2.4.1 shows that 25% or 34 out of our 135 sample banks possess more than two trading memberships at the major exchanges considered in our analysis. However, these banks represent roughly 50% of the number of loans granted or 56% of the loan volume in our sample. We observe banks based in 21 countries, most of

Table 2.4.1: Bank Characteristics

Notes: In this table, we show the summary statistics of the banks' characteristics for our 1,603 bank-year observations. The sample consists of 135 individual banks from 21 countries, examining the period 2003 to 2016. Annual data for all banks is obtained from Standard & Poor's Compustat database. All the characteristics are converted from local currency to USD using the unweighted average of the daily exchange rates in the relevant year. Daily foreign currency exchange rates are obtained from Compustat. Total Assets is the book value of total assets. Trading Securities is USD volume of all trading and dealing accounts divided by total assets. Total Loans and Total Deposits are the book values of all loans granted to non-bank clients divided by total assets and all deposits received from non-bank clients divided by total assets, respectively. Accordingly, the Loans-to-Deposits Ratio is defined as the ratio of Total Loans to Total Deposits. The Capital Ratio is the ratio of the book value of the stockholders' equity to the book value of total assets. The Liquidity Ratio is computed as cash/total assets. # Trading Memberships counts the number of trading memberships in the ten largest security exchanges worldwide and is measured by market volume. Trading Memberships is an indicator variable that is equal to one if a bank has at least one membership and zero otherwise. # Trading Memberships > 2 is an indicator variable that is equal to one if a bank has more than two trading memberships.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
log(Total Assets)	12.491	1.286	10.632	11.306	12.435	13.600	14.898
Trading Securities	0.087	0.116	0.000	0.003	0.033	0.130	0.677
ROA (in %)	0.582	0.543	−1.146	0.275	0.564	0.947	1.760
Capital Ratio	0.066	0.028	0.018	0.045	0.061	0.083	0.133
Liquidity Ratio	0.038	0.033	0.005	0.014	0.027	0.052	0.125
Total Loans	0.498	0.186	0.000	0.402	0.535	0.643	0.719
Total Deposits	0.575	0.240	0.000	0.419	0.616	0.767	0.882
Loans-to-Deposits	0.892	0.414	0.000	0.687	0.803	1.074	1.949
# Trading Memberships	2.025	2.865	0	0	1	2.8	10
Trading Memberships	0.628			–			
# Trading Memberships > 2	0.250			–			

Table 2.4.2: Banks by Region/Country

Notes: In the table, we present the number of banks per country/region and the corresponding mean values within a relevant country/region. For the larger regions (the US, other Advanced economies, and emerging economies), we also report the standard deviation within regions in parentheses. log(Total Assets) is the logarithm of the book value of total assets. # Trading Memberships counts the number of trading memberships in the ten largest security exchanges worldwide measured by market volume. Δlog(Loan Volume) is the year-on-year difference of the logarithm of the loan volume.

Country/Region	# of Banks	log(Total Assets)	# Trading Memberships	Δlog(Loan Volume)
US	41	13.071	3.789	1.803
		(1.273)	(4.564)	(2.476)
Other Advanced	71	13.572	4.004	1.296
		(1.064)	(3.039)	(2.150)
Canada	5	13.045	3.554	1.678
European Union	32	13.868	4.833	1.249
Switzerland	3	13.925	9.939	1.504
Japan	23	13.108	1.308	1.263
Australia	7	12.777	0.993	0.876
Emerging	24	12.857	1.073	0.568
		(1.288)	(1.117)	(1.421)
China	13	13.992	1.885	0.627
Hong Kong	2	11.397	1.000	0.445
Singapore	2	12.084	0.000	0.591
South Korea	4	12.294	0.000	0.311
Taiwan	2	11.352	0.192	0.426
Brazil	1	12.554	1.000	0.586

them located in North America, Europe, and Japan. In Table 2.4.2, we show the distribution of banks across countries and regions. Banks are similar in size and lending across regions. Banks in Europe and North America are particularly active in securities trading, with large numbers of exchange memberships. We observe a total of 8,242 firms, most of which are based in the US, the EU, and Japan. Table 2.4.3 shows a more detailed distribution of firms across countries. Roughly two-thirds of the firms in our sample are based in advanced economies and roughly one-third are based in emerging economies. We show the time series of the Financial Stress Index for the US, other advanced economies, and emerging economies in Figure 2.4.1. "Other advanced economies" comprises primarily Europe and Japan. An index value of zero suggests that stress is at normal levels, positive

Table 2.4.3: Distribution of Borrowers by Country

Notes: In this table, we present the distribution of borrower firms by country in our sample. The data for all firms is obtained from Standard & Poor's Compustat database. We observe a total of 8,242 firms from 81 countries between 2003 and 2016.

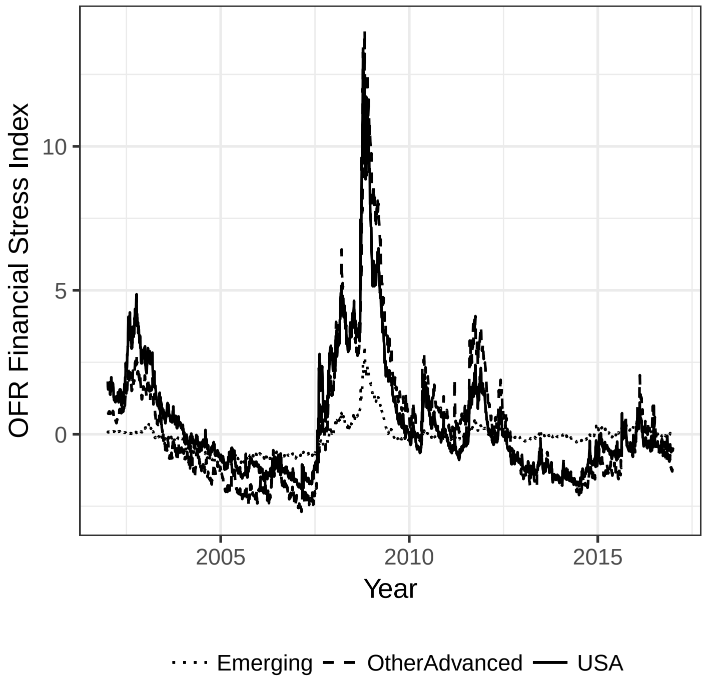
Rank	Country	Number of Firms
1	US	3,755
2	Japan	1,264
3	Canada	480
4	United Kingdom	425
5	Taiwan	269
6	Australia	220
7	France	186
8	Germany	174
9	India	135
10	Hong Kong	115
11	China	89
12	Italy	85
13	South Korea	79
14	The Netherlands	73
15	Spain	68
16	Singapore	59
17	Sweden	56
18	Switzerland	54
19	Norway	53
20	Russia	44
21	Finland	42
22	Malaysia	36
23	Brazil	35
24	New Zealand	34
25	Mexico	31
26 - 81	Others	381
	Total	8,242

(negative) value indicate increased (decreased) stress. USA covers the US economy. The index clearly identifies the last financial crisis for all three regions, indicating extreme financial stress during that time period. In line with previous research the index clearly shows less financial stress in emerging economies compared to advanced economies (see, e.g., Blanchard et al. (2010), Goldstein and Xie (2009), Keat (2009)). The time series for the US and all other advanced economies almost completely overlap during the entire sample period. Both of these time series share a correlation coefficient of roughly 0.94, indicating almost perfect co-movement. The correlation coefficient between the time series for the US and the emerging economies is 0.77 and for the EU and the emerging economies, the correlation coefficient is 0.84. This suggests that a simple crisis dummy would be sufficient to capture the crisis timing globally, but not the severity of the impact. In terms of the severity of the impact of the crisis, it appears to be sufficient to distinguish between advanced and emerging economies. We now turn to the relevant firm clusters in the estimation of model (2.2). The firm clusters are formed based on 8,242 individual firms using the procedure outlined in the beginning of Section 3.4. We form, in total, 1,732 individual clusters, such that the average firm cluster consists of roughly five firms, each firm incorporated in the same country, active in the same industry, and within the same EBITDA interest coverage range. We show the summary statistics for the firm clusters in our sample in Table 2.4.4. The firm clusters in our sample are comparable but rather large in terms of the book value of total assets. However, the clusters are diverse in their leverage and their changes in cash holdings, with the net debt to assets ratios ranging from 4.3% to 68.2% and changes in cash as a share of assets ranging from -5.3% to $+8.4\%$. This clearly indicates a variation in the need of bank financing across our firm clusters.

The bottom three rows in Table 2.4.4 show the summary statistics for our measures for the exposure of firm clusters towards the trading expertise and financial market conditions of their respective lenders, as defined in Equations (2.3) – (2.5). The mean value of 0.45 for $\overline{\textit{Trading Exposure}}$ implies that the average firm cluster in our sample receives 45% of its loans from banks with trading expertise. However, the degree of dependence varies significantly across firm clusters, with some clusters receiving none of their loans from banks with trading expertise and others receiving all of their loans from banks with trading expertise. $\overline{\textit{Trading Bank FSI Exposure}}$ follows the same idea but also takes the level of financial stress in the relevant lender country into account. Thus, for each firm cluster, it

Figure 2.4.1: Is Financial Stress in Crisis Periods the Same Around the World?

Notes: In this figure, we show the value of the Financial Stress Index of the US Office of Financial Research. The index is a measure of systemic financial stress, capturing contributions to financial stress from credit, equity valuations, funding, safe assets, and volatility. An index value of zero suggests that stress is at normal levels, and a positive (negative) value indicates increased (decreased) stress. Other Advanced covers advanced economies other than the US, primarily the EU and Japan. Emerging covers emerging markets. For details on the index computation and coverage, see Monin (2017).



represents a weighted average of the Financial Stress Index across the countries of the cluster's lenders with trading expertise, whereas the weights are the respective lending shares. $\overline{\textit{Bank FSI Exposure}}$ is constructed in the same way but does not distinguish between lenders with and without trading expertise. Note that both measures can be negative or below -1 or $+1$, as the Financial Stress Index is not restricted in its range.

Table 2.4.4: Characteristics of Firm Clusters

Notes: In this table, we show the summary statistics for the firm clusters in our sample. The sample consists of 1,732 individual firm clusters from 81 countries, examining the period 2003 to 2016. The firm clusters are based on 8,242 individual firms, such that the average cluster consists of 4.76 firms. The clusters are formed by matching firms according to their (1) country, (2) industry, and (3) and EBITDA interest coverage, following the approach in Acharya et al. (2018). Data for all firms and foreign currency exchange rates is obtained from Standard & Poor's Compustat database. All non-USD values are converted to USD before any computations. log(Assets) is the logarithm of the book value of total assets. Capex refers to the capital expenditure. Employment Growth is the year-to-year change in the logarithm of the number of employees. Cash includes cash and cash equivalents. Net Debt is the sum of short-term and long-term debt minus cash and cash equivalents. Short-term Debt is all debt with a remaining time to maturity of up to one year, and Long-term Debt is all debt with a remaining time to maturity of more than one year. $\overline{\text{Trading Exposure}}$, $\overline{\text{Trading Bank FSI Exposure}}$, $\overline{\text{Bank FSI Exposure}}$ are defined in Equations (2.3) – (2.5).

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Capex/Assets	0.049	0.035	0.006	0.023	0.040	0.067	0.129
Employment Growth	0.028	0.155	−0.297	−0.041	0.015	0.087	0.387
log(Assets)	8.726	1.672	6.686	7.230	8.524	9.899	12.723
ΔCash/Assets	0.008	0.036	−0.053	−0.013	0.004	0.027	0.084
Net Debt/Assets	0.397	0.176	0.043	0.285	0.410	0.528	0.682
Intangible/Assets	0.161	0.165	0.005	0.023	0.097	0.261	0.551
Ebitda/Assets	0.008	0.009	0.00004	0.001	0.004	0.012	0.027
$\overline{\text{Trading Exposure}}$	0.450	0.300	0.000	0.191	0.510	0.650	1.000
$\overline{\text{Trading Bank FSI Exposure}}$	0.080	1.070	−1.910	−0.545	0.000	1.700	5.560
$\overline{\text{Bank FSI Exposure}}$	0.17	1.84	−1.910	−1.093	−0.090	0.699	5.560

2.5 Results

2.5.1 Credit Supply

In this section, we present the results of the estimation of our model (2.1). The aim of this analysis is to test the hypothesis that banks with trading expertise supply fewer loans to non-financial firms than non-trading banks. We show the estimation results for model (2.1) in Table 2.5.1. Columns (1) to (3) show the results of our three trading expertise proxies but without the interaction with financial crisis measures. When proxying for trading expertise through a dummy that equals one if a bank has at least one trading membership at a securities exchange as in column (1), we do not find a statistically significant effect on credit supply. However, for both *# Trading Memberships* and *# Trading Memberships* > 2, we find a negative, statistically significant effect. This supports our hypothesis that banks' trading activity negatively affects credit supply. The effect of *# Trading Memberships* is smaller than that of *# Trading Memberships* > 2. The effects also differ largely in their economic significance. This lends support to our argument that a small number of trading memberships does not necessarily indicate trading expertise in large banking organizations. The coefficient associated with *# Trading Memberships* indicates an average reduction in loan supply by approximately 2.66% per additional trading membership. The coefficient associated with *# Trading Memberships* > 2 indicates a reduction in the credit supply of approximately 19.18%. Thus, both effects are not only statistically significant, but also economically meaningful.

We draw three conclusions from these results. First, the finding that the simple *Trading Membership* dummy turns out to be insignificant while the *# Trading Memberships* > 2 dummy turns out to be highly significant supports our earlier assertion that, especially for large banks, a single or a small number of trading memberships does not indicate specific trading expertise compared to other, similar large banks. A small number of trading memberships may indeed be purchased to service clients' trading needs or the banks' hedging needs rather than engaging in proprietary trading.[12]

[12] It is worth noting that we differ in this regard from Abbassi et al. (2016), who only use a dummy variable at the largest fixed-income securities trading platform in the German market. Their sample of German banks is by far more heterogeneous in terms of bank size compared to our sample. Thus, in the case of Abbassi et al. (2016), for a comparatively small bank, a single trading membership may already indicate greater trading expertise, as the business model differs from the business model of the large banks in our sample.

Second, our dummy variables *Trading Memberships* and *# Trading Memberships* > 2 are constructed such that the *# Trading Memberships* > 2 banks are a subset of the *Trading Membership* banks.[13] This finding suggests non-linearity in the relationship between changes in credit supply and the number of trading memberships. Such non-linearity also explains the rather large difference in magnitude between the effects of *# Trading Memberships* and *# Trading Memberships* > 2, as *# Trading Memberships* implies a linear relationship. For this reason, we proceed with the *# Trading Memberships* > 2 dummy, i.e., the specification in column (3), as our baseline model. Finally, we conclude that our evidence as presented in Table 2.5.1 columns (1) to (3) supports the hypothesis that banks with greater trading expertise reduce credit supply.

Columns (4) to (6) contain the estimation results for the specifications, that include interactions between the *# Trading Memberships* > 2 dummy and the financial crisis measures, as shown in model (2.1). In column (4), we interact the *# Trading Memberships* > 2 dummy with the Financial Stress Index. Since the index is bank country-specific and time-varying we must drop the bank country-year fixed effects and include bank country fixed effects instead. The coefficient associated with *# Trading Memberships* > 2 now captures the effect of trading expertise for Financial Stress Index values of zero, i.e. in absence of either positive or negative stress. The negative and statistically significant coefficient associated with the interaction term indicates that banks with trading expertise tend to reduce their credit supply by an additional 3.25% per unit increase in financial stress compared to the 19% baseline reduction (statistics are approximate). In contrast, we do not find a significant effect for the Financial Stress Index alone. While evidently there was a great impact of the 2007 to 2009 financial crisis on the corporate loan market, this could suggest that the Financial Stress Index can capture the direct link between crisis and lending, but only through banks' securities trading activities.

In Figure 2.5.1, we visualize the marginal effect of *# Trading Memberships* > 2 for the observed range of values of the Financial Stress Index.[14]. The positive values of the index indicate financial market stress, and negative values indicate financial market relaxation (stabilizing conditions). The marginal effect is a downward slope, statistically significant, and negative, with a relatively narrow confidence interval across the whole range of

[13] The *Trading Membership* dummy could also be called *# Trading Membership* > 0.

[14] See Figure 2.4.1

Financial Stress Index values. This supports our earlier interpretation that banks with trading expertise tend to reduce credit supply even under favorable financial market conditions but reduce credit supply even further with increasing financial market stress. In the figure, we highlight the marginal effects for the zero values of the Financial Stress Index, its 2007 to 2009 crisis average, and its 2007 to 2009 crisis peak value. The corresponding economic effects are approximately 19%, 33%, and 47.8%, respectively.

The Financial Stress Index is not a crisis indicator in this sense. Only large values indicate a financial crisis, while the index fluctuates around zero throughout the business cycle. Thus, to more clearly isolate the impact of the financial crisis in column (6) of Table 2.5.1, we substitute the Financial Stress Index with a simple dummy variable that equals one during the crisis 2007 to 2009 and zero otherwise. Consequently, we cannot include either the firm cluster-year or bank country-year fixed effects in this specification. We instead use firm cluster and bank country fixed effects. This dummy can capture the effects of the financial crisis on credit supply more broadly and more directly than the Financial Stress Index. All effects in column (6) are negative and statistically significant. The results indicate that banks with trading expertise reduced their loan supply by approximately 46.23% during the financial crisis while banks without trading expertise reduced their loan supply only by approximately 29.69%. The results in columns (4) and (6) support the hypothesis that banks with trading expertise reduce their loan supply more during periods of financial turmoil. However, the results in column (6) should be taken lightly, as the adjusted R^2 drops by almost half compared to the other specification. While the adjusted R^2 still indicates that our model is reasonably useful, it also suggests the large importance of the firm cluster-year fixed effects, which proxy for credit demand. Finally, in column (5), we present the results of the interaction between our bank-specific capital market index and the *# Trading Memberships* > 2 dummy. Neither the interaction nor the capital markets index are statistically significant. The index can proxy for a bank's exposure to price fluctuation in the securities in which the bank is actually investing. However, it seems that our index cannot capture this phenomenon well enough.

In summary, we find support for the hypothesis that banks with greater trading expertise provide fewer loans to non-financial borrowers and reduce credit supply specifically during financial crises. Both findings are consistent with the theoretical predictions of Diamond and Rajan (2011), Shleifer and Vishny (2010), and Boot and Ratnovski (2016).

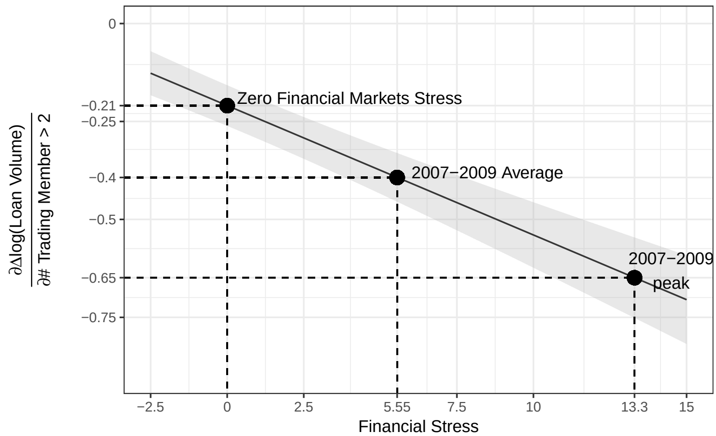
Table 2.5.1: Trading Expertise and Bank Lending

Notes: In this table, we present the results of a modified version of the Khwaja and Mian (2008) model. The unit of observation is firm cluster-bank-year. Firm clusters are formed based on a firm's country of incorporation, the two-digit SIC code, and a firm's credit rating, estimated based on the median EBIT interest coverage ratios. Trading Membership equals one if a bank has at least one trading membership at a major stock exchange and zero otherwise. # Trading Memberships represents the number of a bank's trading memberships at major stock exchanges. # Trading Memberships > 2 equals one if # Trading Memberships is greater than two and zero otherwise. Capital Markets is the market capitalization weighted average over the change Standard & Poor's Global Equity Index for the countries in which a bank possesses a trading membership at the regional stock exchange and the bank's country of incorporation. Financial stress is the value of the Financial Stress Indicator as provided by the US OFR for a bank's country of incorporation. All regressions include bank-level controls (the logarithm of total assets, return-on-assets, common equity/total assets, cash/total assets, total loans/total deposits). Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Δlog(Loan Volume)					
	(1)	(2)	(3)	(4)	(5)	(6)
Trading Membership	0.020					
	(0.029)					
# Trading Memberships		-0.027^{***}				
		(0.004)				
# Trading Memberships > 2			-0.211^{***}	-0.209^{***}	-0.214^{***}	-0.160^{***}
			(0.027)	(0.027)	(0.027)	(0.026)
Financial Stress$_t$				0.007		
				(0.008)		
Capital Markets$_t$					0.036	
					(0.075)	
Crisis						-0.220^{***}
						(0.012)
(# Trading Memberships > 2)*Financial Stress$_t$				-0.033^{***}		
				(0.004)		
(# Trading Memberships > 2)*Capital Markets$_t$					0.021	
					(0.045)	
(# Trading Memberships > 2)*Crisis						-0.256^{***}
						(0.018)
Observations	268,910	268,910	268,910	268,910	268,910	268,910
Adjusted R^2	0.374	0.374	0.374	0.371	0.374	0.192
Bank Controls	YES	YES	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES	YES	NO
Bank Country-Year FE	YES	YES	YES	NO	YES	NO
Firm Cluster FE	NO	NO	NO	NO	NO	YES
Bank Country	NO	NO	NO	YES	NO	YES

Figure 2.5.1: The Effect of Trading Expertise as a Function of Financial Stress

Notes: In this figure, we show the marginal effect of the # Trading Memberships > 2 dummy for different levels of the Financial Stress Index published by the US OFR. The Financial Stress Index is centered on zero. The positive values of the index indicate financial market stress, and negative values indicate financial market relaxation (stabilizing conditions). See Figure 2.4.1 for a plot of the time series of the index. The axis below shows the observed index values during our sample period. During the 2007 to 2009 financial crisis, the index peaked, close to a value of 15, with an average across 2007 to 2009 of approximately 5.55. The effect is computed using the coefficients reported in column (4) of Table 2.5.1. The solid bold line represents the marginal effect, and the shaded area represents the corresponding 95% confidence interval using cluster robust standard errors.



2.5.2 Real Effects

In this section, we investigate whether the reduced credit supply of banks with greater trading expertise translates into lower capital expenditure or lower employment growth of their borrowers. We use model (2.2) to test this channel of spillover from the banking sector to the real economy. Table 2.5.2 shows the results for capital expenditure as the dependent variable. In columns (1) and (2), we consider a direct transmission channel of banks' securities trading to capital expenditure through lending. $\overline{\textit{Trading Exposure}_t}$ measures the share of loans granted to a firm cluster from banks with more than two trading memberships, i.e., those with trading expertise, in year *t*. In column (1), we only include $\overline{\textit{Trading Exposure}_t}$, while in column (2), we interact this measure with a 2007 to 2009 financial crisis dummy.[15] The crisis dummy itself turns out to be positive and significant. While surprising at first, this can be explained by the delayed response of capital expenditure to the financial crisis, which cannot be captured by the dummy as it does not allow for distinction between crisis years. Figure 2.5.2 a) shows that capital expenditure continued to increase (albeit at a lower rate) at the onset of the financial crisis from 2007 to 2008. Only then did capital expenditure decline sharply, reaching its minimum in 2010 when the situation in the financial sector had already normalized.[16] The interaction term in column (2) is statistically significant, indicating that there is a negative effect of exposure to banks with trading expertise during the financial crisis. In column (3), we repeat the model from column (2) but now use the Financial Stress Index instead of a simple crisis dummy to gain a more granular measurement of the impact of the crisis. Using the Financial Stress Index rather than a crisis dummy does not only allow us to distinguish between different crisis years, but also allows for the distinction of differences in the intensity of the financial crisis in different countries. The disadvantage of using the Financial Stress Index is its narrow focus on capital market conditions. Thus, the index does not capture all aspects of the crisis. The Financial Stress Index is negative and statistically significant and thus has the expected sign. However, the interaction term between the Financial Stress Index and $\overline{\textit{Trading Exposure}_t}$ is insignificant. This does not necessarily contradict the results reported in column (2). While the crisis dummy in column (2) captures a globally uniform impact

[15]To be able to also include year fixed effects in this specification, we need to drop one additional year dummy.

[16]This behavior is consistent with the results reported in Kahle and Stulz (2013).

of the crisis, the Financial Stress Index here is specific to capital market stress in the firm cluster's country of incorporation. Thus, the insignificant interaction term merely implies that increased stress in a borrower's home country does not affect capital expenditure. Note that our firm clusters consist of rather large companies who are listed on stock exchanges. Smaller companies might be affected by local capital markets conditions to a greater extent.

Next, we investigate whether financial market stress in lender countries affects capital expenditure. In column (4), we include $\overline{\textit{Bank FSI Exposure}}_t$, which is simply a weighted average of the values of the Financial Stress Indices of all lenders of a firm cluster, weighted by the respective share of the total loans this firm cluster received. Thus, we test whether there is a transmission of financial stress from banks to their borrowers unconditional on whether the banks have trading expertise or not. The coefficient associated with $\overline{\textit{Bank FSI Exposure}}_t$ is statistically insignificant, indicating no such effect. Finally, in column (5), we investigate whether there is a transmission of financial market stress to borrowers' capital expenditure by banks with trading expertise. We include $\overline{\textit{Trading Bank FSI Exposure}}_t$, which has the same interpretation as $\overline{\textit{Bank FSI Exposure}}_t$ but only takes banks with trading expertise into account. Now, the coefficient is statistically significant and negative, indicating that borrowers tend to reduce their capital expenditure if their dependence on banks with trading expertise increases and if these banks are experiencing financial stress in their home markets. The results concerning employment growth in Table 2.5.3 indicate that the effect of $\overline{\textit{Bank FSI Exposure}}_t$ is negative and highly statistically significant. Moreover, the coefficients of $\overline{\textit{Trading Exposure}}_t$ and its interaction with our crisis dummy variable in column (2) are statistically insignificant. However, to get a better understanding of the effect of $\overline{\textit{Trading Exposure}}_t$ during the 2007 to 2009, crisis we plot the marginal effect of the crisis dummy on employment growth in Figure 2.5.3. We plot the marginal effect of the crisis dummy rather than the marginal effect of $\overline{\textit{Trading Exposure}}_t$, since we have from Equation (2.2): $\partial\text{Employment Growth}/\partial\text{Crisis} = \Psi_{\textit{Crisis}} + \lambda \times \overline{\textit{Trading Exposure}}$. Thus, we investigate the effect of the crisis on employment growth for the various levels of $\overline{\textit{Trading Exposure}}_t$. The figure shows the negative and statistically significant effect of the crisis on employment growth that becomes stronger as $\overline{\textit{Trading Exposure}}_t$ increases. However, the additional decrease in employment growth due to $\overline{\textit{Trading Exposure}}_t$ is limited. While firms that do not borrow from banks with trading expertise reduce the number of employees by

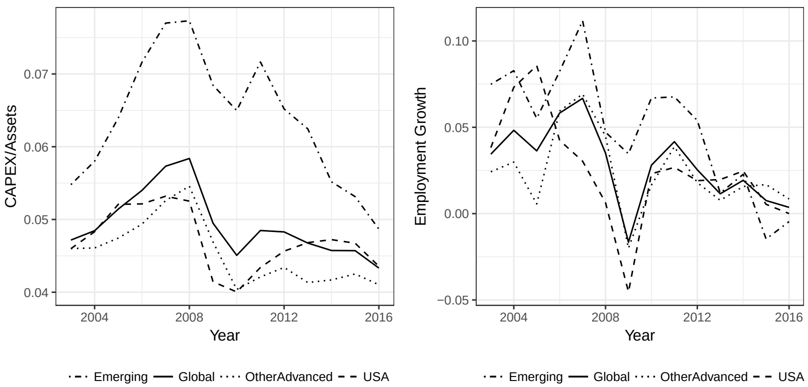
Table 2.5.2: Does Trading or Crisis Exposure Affect Capex?

Notes: In the table we present the results of the firm cluster level regressions. The unit of observation is firm cluster-year. The dependent variable is capital expenditure (Capex). The exposure measures are defined as in Equations (2.3) to (2.5). Crisis is a dummy variable equal to one for the years 2007 to 2009. If Crisis is included in the regression, the respective year fixed effects are dropped. FSI refers to the value Financial Stress Indicator, as developed by the US Office for Financial Research for the firm's country of incorporation. $\overline{\text{Bank FSI Exposure}}_t$ measures the firm's exposure to the FSI values of its lending banks, while $\overline{\text{Trading Bank FSI Exposure}}_t$ measures exposure to the FSI values of lending banks that also possess trading expertise. All regressions include country, industry, and year fixed effects and one-year lagged firm cluster control variables (the logarithm of total assets, net debt/total assets, intangible assets/total assets, ΔCash/total assets, and EBITDA/total assets). Information on firm-bank lending relationships is taken from Thomson Reuter's LPC DealScan database. Firm data is obtained from Standard & Poor's Compustat database. Our samples ranges from 2003 to 2016. All values are transformed to USD using the appropriate foreign exchange rates from Compustat. All standard errors are clustered at the firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: $Capex_t$/Total $Assets_t$				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{Trading Exposure}}_t$	0.001	0.002	0.001		
	(0.002)	(0.002)	(0.002)		
Crisis		0.006^{***}			
		(0.001)			
Financial Stress_t			-0.001^{***}	-0.001^{***}	-0.001^{***}
			(0.0004)	(0.0004)	(0.0004)
$\overline{\text{Trading Exposure}}_t$*Crisis		-0.003^{*}			
		(0.002)			
$\overline{\text{Trading Exposure}}_t$*Financial Stress_t			-0.0003		
			(0.0004)		
$\overline{\text{Bank FSI Exposure}}_t$				-0.0002	
				(0.001)	
$\overline{\text{Trading Bank FSI Exposure}}_t$					-0.001^{*}
					(0.0004)
Observations	17,768	17,768	17,768	17,768	17,768
Adjusted R^2	0.385	0.386	0.386	0.386	0.386
Country FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Firm Cluster Controls	YES	YES	YES	YES	YES

Figure 2.5.2: Capex and Employment Growth Development Over Time

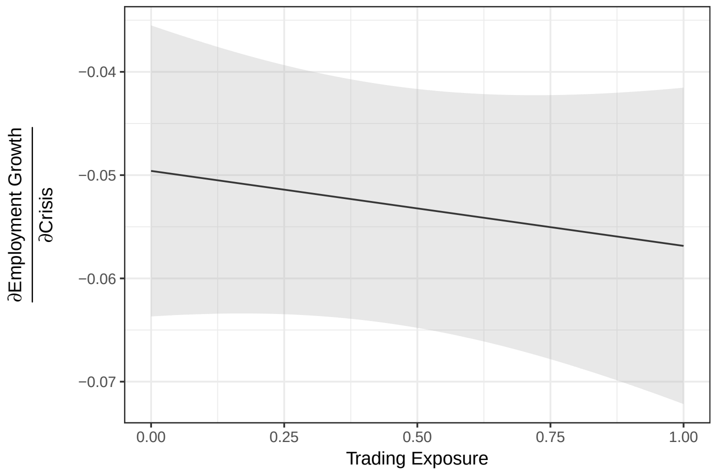
Notes: In panels a) and b) below, we present the time series of cross-sectional mean values of capital expenditure as a share of total assets and employment growth across regions and globally for firm clusters incorporated in the US, other advanced economies, and emerging economies as defined by the US Office of Financial Research. Employment growth at the firm cluster level is computed as the year-to-year change in the logarithm of the number of employees.



a) Capital Expenditure (Capex) b) Employment Growth

Figure 2.5.3: The Marginal Effect of Trading Exposure During a Crisis

Notes: In this figure, we show the marginal effect of the 2007 to 2009 crisis dummy for different levels of the variable $\overline{\text{Trading Exposure}}_t$. The effect is computed using the coefficients reported in column (2) of Table 2.5.3, assuming that the crisis dummy is set to one. On the x-axis we show the range of values of $\overline{\text{Trading Exposure}}_t$ in our sample. The solid bold line represents the marginal effect. The shaded area represent the 95% confidence interval around the marginal effect using cluster robust standard errors.



approximately 4.8%, firms that obtain all of their borrowing from banks with trading expertise reduce the number of employees by approximately 5.5%.

In summary, we find that dependency on banks with trading expertise does not only negatively affect borrowers' investments in capital but also affects employment growth.

Table 2.5.3: Are Trading or Crisis Exposure Affecting Employment Growth?

Notes: In this table, we present the results of the firm cluster level regressions. The unit of observation is firm cluster-year. The dependent variable is employment growth, measured as the year-to-year change in the logarithm of the number of employees. The exposure measures are defined in Equations (2.3) to (2.5). Crisis is the dummy variable, equal to one for the years 2007 to 2009. If Crisis is included in the regression, the respective year fixed effects are dropped. FSI refers to the value Financial Stress Indicator, as developed by the US Office for Financial Research, for the firm's country of incorporation. $\overline{\text{Bank FSI Exposure}_t}$ *measures the firm's exposure to the FSI values of its lending banks, while* $\overline{\text{Trading Bank FSI Exposure}_t}$ *measures exposure to the FSI values of lending banks that also possess trading expertise. All regressions include country, industry, and year fixed effects and one year lagged firm cluster control variables (the logarithm of total assets, net debt/total assets, intangible assets/total assets, ΔCash/total assets, and EBITDA/total assets). Information on firm-bank lending relationships is taken from Thomson Reuters LPC DealScan database. Firm data is obtained from Standard & Poor's Compustat database. Our sample ranges from 2003 to 2016. All values are converted to USD using the appropriate foreign exchange rates from the Compustat database. All standard errors are clustered at the firm cluster level. Significance levels:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: Employment Growth$_t$				
	(1)	(2)	(3)	(4)	(5)
$\overline{\text{Trading Exposure}_t}$	-0.012^{*}	-0.009	-0.012^{*}		
	(0.006)	(0.007)	(0.006)		
Crisis		-0.005			
		(0.007)			
Financial Stress$_t$			-0.001	0.001	-0.001
			(0.002)	(0.002)	(0.002)
$\overline{\text{Trading Exposure}_t}$*Crisis		-0.011			
		(0.009)			
$\overline{\text{Bank FSI Exposure}_t}$				-0.008^{***}	
				(0.002)	
$\overline{\text{Trading Bank FSI Exposure}_t}$					-0.001
					(0.002)
Observations	17,768	17,768	17,768	17,768	17,768
Adjusted R^2	0.060	0.060	0.060	0.061	0.060
Country FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Firm Cluster Controls	YES	YES	YES	YES	YES

2.6 Further Analysis

2.6.1 Loan Pricing

In this section, we present some additional results regarding the differences in loan pricing between banks with higher and lower trading expertise. Following the same approach as Acharya et al. (2018), we analyze loan pricing simply by adapting our Khwaja and Mian (2008) estimator in Equation (2.1) to the change in loan prices rather than the change in loan volume. Note that the Khwaja and Mian (2008) estimator and specifically the argument that the included borrower-year fixed effects can capture variations in unobserved credit demand is derived from a microeconomic model that relies on the loan volume.[17] However, prices are also driven by demand. Thus, we believe that our empirical specification in Equation (2.1) remains valid if we use the change in loan prices as dependent variable.

We measure loan prices as All-in Spread drawn, which equals the total (fees and interest) annual spread paid over LIBOR drawn from the loan. In particular, we calculate All-in Spread drawn = Upfront fee + Annual fee + Utilization Fee + Interest Spread over LIBOR. Thus, the All-in Spread drawn represents the cost of financing for the relevant borrower (see, e.g., Ivashina (2005)).

We present the results of this analysis in Table 2.6.1. Our results indicate that all of our sample banks increase loan prices during periods of financial stress. Moreover, we find that banks with more trading expertise (measured either through *# Trading Memberships* > 2 or *# Trading Memberships*) charge their borrowers higher prices for drawn loans. However, the interaction between the two effects is statistically insignificant. This indicates that while banks with more trading expertise generally tend to charge higher prices, they do not behave differently in terms of their loan pricing than banks without trading expertise during periods of crisis.

2.6.2 Are US Banks Different?

In our sample, the US is the most common country of origin of the banks. Thus, while we have a global sample, naturally, the question arises concerning to what extent our results are

[17] See Khwaja and Mian (2008) for details regarding the derivation.

Table 2.6.1: The Effect of Trading on Loan Pricing

Notes: In this table, we present the results regarding the effect of trading expertise on loan pricing. The unit of observation is firm cluster-year. The dependent variable is the change of the logarithm in the All-in Spread Drawn, while we calculate All-in Spread drawn = Upfront fee + Annual fee + Utilization Fee + Interest Spread over LIBOR. Firm clusters are formed based on a firm's country of incorporation, the two-digit SIC code, and a firm's credit rating, estimated based on the median EBIT interest coverage ratios. # Trading Memberships represents the number of a bank's trading memberships at major stock exchanges. # Trading Memberships > 2 equals one if # Trading Memberships is greater than two and zero otherwise. Financial Stress is the value of the Financial Stress Indicator, as provided by the US OFR, for a bank's country of incorporation. All regressions include bank-level controls (the logarithm of total assets, return-on-assets, common equity/total assets, cash/total assets, and total loans/total deposits). Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: Δ(All-in Spread Drawn)			
	(1)	(2)	(3)	(4)
# Trading Memberships > 2	0.757^{***}		0.781^{***}	
	(0.196)		(0.198)	
# Trading Memberships		0.088^{***}		0.092^{***}
		(0.026)		(0.027)
Financial Stress	0.477^{**}	0.478^{**}	0.572^{**}	0.580^{***}
	(0.212)	(0.212)	(0.226)	(0.225)
(# Trading Memberships > 2)*Financial Stress			−0.129	
			(0.125)	
(# Trading Memberships)*Financial Stress				−0.022
				(0.019)
Observations	203,947	203,947	203,947	203,947
Adjusted R^2	0.266	0.266	0.266	0.266
Bank Controls	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES
Bank Country	YES	YES	YES	YES

driven by US banks. In particular, these US banks include investment banking giants Goldman Sachs, Morgan Stanley, J.P. Morgan Chase, and Bank of America Merrill Lynch. We investigate whether there is a difference between US and non-US banks by repeating our analysis of model (2.1) for the US and non-US bank sub-samples. We show the results of this exercise in Table 2.6.2. Columns (1) and (2) present the results concerning the US sub-sample, while columns (3) and (4) present the results concerning the non-US sub-sample. A comparison of columns (1) and (3) shows that the trading expertise dummy *# Trading Memberships* > 2 is negative and statistically significant for US banks but not for non-US banks, while both coefficients are also statistically different from each other. This indicates that US banks with trading expertise indeed behave differently from non-US banks with trading expertise.

Next, a comparison of the interaction terms between trading expertise and financial stress in columns (2) and (4) shows that the interactions are negative and statistically significant in both sub-samples, while trading expertise in itself remains only significant in the US sub-sample. Furthermore, the levels of the coefficients associated with the Financial Stress Index and with the interaction term between the Financial Stress Index and *# Trading Memberships* > 2 differ between columns (2) and (4). Non-US banks without trading expertise reduce their credit supply by an approximate 4.3% per unit increase in the Financial Stress Index, while US banks without trading expertise reduce their credit supply by an approximate 11% per unit increase in the Financial Stress Index. Furthermore, non-US banks with trading expertise reduce their credit supply by an approximate 8% per unit increase in the Financial Stress Index, while we estimate that US banks with trading expertise reduce their credit supply by an approximate 13% per unit increase in the Financial Stress Index. These results demonstrate the great sensitivity of US banks to financial market stress, regardless of the banks' trading expertise. Comparing regression results across sub-samples can be problematic. Therefore, we also re-estimate our (2.1) but include a dummy variable that equals one for banks headquartered in the US and zero otherwise. We interact this US banks dummy with all variables in (2.1). To account for differences in regulatory and macroeconomic environments among countries, we further augment the model with a set of bank country dummy variables for countries other than the US.[18] This allows us to estimate different slope coefficients for US and non-US bank

[18]In other words, we include a full set of bank country dummy variables, but we only interact the US

Table 2.6.2: Are US Banks Different? – A Sub-Sample Analysis

Notes: In this table, we present the results regarding the effect of trading expertise on credit supply for the sub-samples that contain only US or non-US banks. The unit of observation is firm cluster-year. Firm clusters are formed based on a firm's country of incorporation, the two-digit SIC code, and a firm's credit rating, estimated based on median EBIT interest coverage ratios. # Trading Memberships represents the number of a bank's trading memberships at major stock exchanges. # Trading Memberships $>$ 2 equals one if # Trading Memberships is greater than two and zero otherwise. Financial Stress is the value of the Financial Stress Indicator, as provided by the US OFR, for a bank's country of incorporation. All regressions include bank-level controls (the logarithm of total assets, return-on-assets, common equity/total assets, and total loans/total deposits). Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: Δlog(Loan Volume)			
	US Banks		non-US Banks	
	(1)	(2)	(3)	(4)
# Trading Memberships $>$ 2	-0.431^{***}	-0.529^{***}	0.019	0.021
	(0.047)	(0.044)	(0.031)	(0.030)
Financial Stress		-0.117^{***}		-0.044^{***}
		(0.006)		(0.008)
(# Trading Memberships $>$ 2)*Financial Stress		-0.022^{**}		-0.037^{***}
		(0.009)		(0.004)
Observations	66,065	66,065	202,845	202,845
Adjusted R^2	0.407	0.272	0.384	0.379
Bank Controls	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES
Bank Country-Year FE	YES	NO	YES	NO
Bank Country	NO	NO	NO	YES
Year FE	YES	NO	NO	NO

with in a single regression equation. We show the results of this exercise in Table 2.6.3. The first three rows (1. - 3.) in the table show the estimated coefficients if the US banks dummy equals zero, while the next three rows (4. - 6.) show the estimated coefficients if the US banks dummy equals one. To assess the statistical significance of the difference in the coefficients we also perform a Wald test for the null hypothesis that the coefficients for US and non-US banks are equal. Thus, we test $H_0 : 1.) = 4.)$ and $H_0 : 1.) = 4.)$ and $3.) = 6.)$ for columns (1) and (2) respectively. The corresponding test statistic is reported at the bottom of the table.

Consistent with our sub-sample analysis, the hypothesis of the equal slope coefficients of US and non-US banks is clearly not supported. While the results of the sub-sample analysis remain unchanged qualitatively, the magnitude of the difference in the coefficients between US and non-US banks becomes larger. To visualize the difference in the behavior of US and non-US banks, we plot the marginal effect of *# Trading Memberships* > 2 for US and non-US banks in Figure 2.6.1. If Financial Stress is zero, non-US banks tend not to change their credit supply, while there is a pronounced negative effect for US banks corresponding to a reduction in credit supply of approximately 45%. As the value of the Financial Stress Index increases, the effect of trading expertise on credit supply increases for both US and non-US banks. However, the total reduction in the credit supply of US banks remains significantly higher as the total reduction in credit supply of non-US banks. In summary, this analysis suggests that our previous results for economically stable times are driven by the behavior of US banks, i.e., it is mainly US banks with trading expertise who cut their credit supply during stable economic times, and they cut their credit supply even further during periods of crisis. However, non-US banks with trading expertise do not reduce their credit supply during economically stable times. They do so during crises. Thus, the behavior of non-US banks is on the one hand consistent with the theoretical prediction that banks with trading expertise reduce their credit supply during crisis to be able to invest in assets for fire-sale prices (see, e.g., Diamond and Rajan (2011), Shleifer and Vishny (2010)). On the other hand, we do not see evidence for the theoretical prediction that non-US banks also reduce their credit supply in economically stable times to allocate funds to (scalable and rather short-term) trading instead of relationship banking activities, such as lending (see Boot and Ratnovski (2016)). However, for US banks, both channels apply, as

banks dummy with the other covariates.

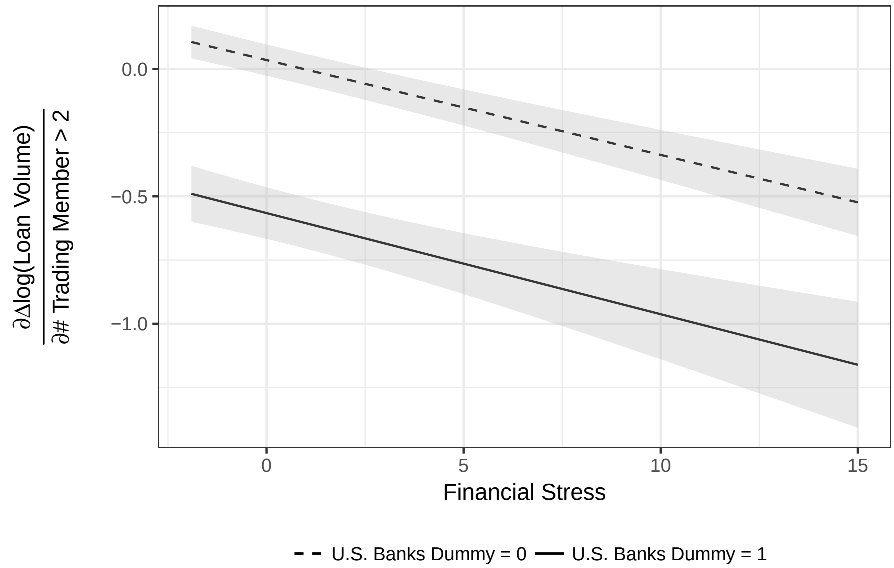
Table 2.6.3: Are US Banks Different? – Single Equation

Notes: In this table, we present the results regarding the effect of trading expertise on credit supply for US or non-US banks. Instead of splitting the sample, we introduce a US banks dummy variable and interact it with the # Trading Memberships > 2 dummy, the Financial Stress Index, and all bank-level controls. The upper part of the table shows the estimated coefficients if the US banks dummy equals zero, and the lower part shows the estimated coefficients if the US banks dummy equals one. The unit of observation is firm cluster-year. Firm clusters are formed based on a firm's country of incorporation, the two-digit SIC code, and a firm's credit rating, estimated based on the median EBIT interest coverage ratios. # Trading Memberships represents the number of a bank's trading memberships at major stock exchanges. # Trading Memberships > 2 equals one if # Trading Memberships is greater than two and zero otherwise. Financial Stress is the value of the Financial Stress Indicator, as provided by the US OFR, for a bank's country of incorporation. All regressions include bank-level controls (the logarithm of total assets, return-on-assets, common equity/total assets, and total loans/total deposits). The Wald statistic corresponds to a test of H_0 : 1.) = 4.) and H_0 : 1.) = 4.) and 3.) = 6.) for columns (1) and (2) respectively. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: Δlog(Loan Volume)	
	(1)	(2)
US Banks Dummy = 0		
1.) # Trading Memberships > 2	0.029	0.035
	(0.031)	(0.031)
2.) Financial Stress		-0.015^{*}
		(0.008)
3.) (# Trading Memberships > 2)*Financial Stress		-0.037^{***}
		(0.004)
US Banks Dummy = 1		
4.) # Trading Memberships > 2	-0.596^{***}	-0.601^{***}
	(0.060)	(0.060)
5.) Financial Stress		-0.017^{**}
		(0.007)
6.) (# Trading Memberships > 2)*Financial Stress		-0.003
		(0.009)
Wald Statistic	121.455	123.598
	(df=1)	(df=2)
Observations	268,910	268,910
Adjusted R^2	0.372	0.373
Bank Controls	YES	YES
Firm Cluster-Year FE	YES	YES
Non-US Bank Country Dummies	YES	YES

Figure 2.6.1: The Marginal Effect of Trading Expertise – A Single Equation

Notes: In this figure, we present the marginal effect of Trading Expertise for different levels of the Financial Stress Index. The effects are based on the regression specification reported in column (2) in Table 2.6.3, setting the # Trading Memberships $>$ 2 dummy equal to one. The dashed line shows the marginal effect if the US banks dummy equals zero, while the solid line shows the marginal effect if the US banks dummy equals one. The shaded areas represent the 95% confidence intervals around the marginal effects using cluster robust standard errors.



US banks with trading expertise reduce their credit supply both in times of crises and in periods of stability.

This difference in behavior between US and non-US banks is important in the assessment of differences in the regulatory regimes regarding proprietary trading by banks in the US and e.g. the UK or the EU. The U.S. Volcker Rule goes further than its UK or EU equivalents as it bans banks from engaging in proprietary trading, while UK and EU regulations only force an insulation of banking activities from trading losses within a banking group without actually banning proprietary trading.

2.6.3 Trading Expertise and Foreign Lending

Home biases in lending and a general decline in foreign lending are well documented in the literature (see, e.g., Marchetti (2016)). In this section, we aim to contribute to this literature by analyzing the effect of banks' trading expertise on foreign lending, and we define foreign lending as loans granted by a bank to a borrower that is incorporated in a country other than the bank.[19] Thus, we create a new variable that equals one if a bank's and borrower's country of incorporation differ and zero otherwise.

To gain further insight, we also create measures of the geographic and economic distance between a bank's and borrower's country of incorporation. We compute the geographic distance between the bank and borrower countries using the great-circle distance formula used in physics and navigation. The great-circle distance is the shortest distance between any two points on the surface of a sphere and is computed as

$$Distance_{i,j} = r \times arccos\left(sin\left(Lat_i\right) sin\left(Lat_j\right) + cos\left(Lat_i\right) cos\left(Lat_j\right) cos\left(Long_i - Long_j\right)\right)$$

where Lat_i, Lat_j and $Long_i$, $Long_j$ are the latitude and longitude respectively, of the centroids of the bank country i and borrower country j.[20] r is Earth's mean radius in km ($\approx$ 6,371 km). Since $Distance_{i,j}$ is heavily skewed, we use the logarithm of $Distance_{i,j}$ in all regressions. If bank and borrower are incorporated in the same country, we set

[19] In the case of syndicated loans, i.e., loans that are granted by multiple banks forming a syndicate, the loan is classified as foreign lending if at least one bank is incorporated in another country.

[20] The centroid of a country is the geometric center of the two-dimensional polygon spanned by the country's borders.

$Distance_{i,j} = 1$ instead of using the great-circle distance.[21] Thus, the logarithm of the geographic distance equals zero whenever the foreign lending dummy equals zero.

We proxy for the economic distance between bank and borrower countries using the absolute value of the difference in the KOF Globalisation Index.[22] The index is a measure of the level of globalization of individual countries along economic, social, and political dimensions.[23]

We repeat our estimation of our model (2.1) but augment the regression model with the foreign lending dummy, the geographic distance measure, and the economic distance measure. In columns (1) and (2) of Table 2.6.5, we show the results for the foreign lending dummy. The coefficient associated with foreign lending is negative and statistically significant, indicating a rather strong lending home bias among the banks in our sample. The coefficient of interest in this regression specification is the interaction between our trading expertise measure *# Trading Memberships* > 2 and the foreign lending dummy. The interaction term coefficients are positive and statistically significant. Thus, banks with trading expertise tend to increase their credit supply to foreign markets by 1.2% to 2.5% compared to banks without trading expertise. At the same time, consistent with our previous results, banks with trading expertise reduce credit supply to their home market by 30% to 32% compared to banks without trading expertise. This behavior changes as the level of financial stress in the banks' home country increases. During a crisis, the banks with trading expertise reduce their credit supply to foreign markets by about 16% compared to banks without trading expertise.[24]

In columns (3) and (4) of Table 2.6.5, we repeat the analysis but use the geographic distance between the bank and borrower countries instead of the foreign lending dummy, and in columns (5) and (6), we use the economic distance.[25] In either case, the conclusions

[21]This approach simply implies that we assume that the physical distance in km between bank and borrower is 1 km if both are incorporated in the same country.

[22]The KOF Index is computed and published by the Swiss Economic Institute at ETH Zürich.

[23]For details regarding the computation of the index, see Dreher (2006) (the original version of the index) and Gygli, Haelg, and Sturm (2018) (the revised version of the index which is used in this paper). Since the most current KOF Globalisation Index is only available until the year 2015, we augment the values for 2016 for each country using simple AR(p) one-year ahead forecasts, while for each country's time-series, the lag-length p is selected to minimize the AIC. Using data until 2015 does not change the results.

[24]This position assumes a level of Financial Stress Index that is equal to 5.55, which corresponds to the average level of financial stress in advanced economies during the 2007 to 2009 financial crisis.

[25]In columns (5) and (6) of Table 2.6.5, the sample size is smaller than in the other regressions as the KOF Globalisation Index is not available for some of the countries in our sample. For example, while we observe

remain the same as for the foreign lending dummy. However, using the geographic distance or economic distance reveals that the effect of banks with trading expertise increasing their loan supply in foreign markets while reducing their credit supply in their home market becomes stronger as the distance between bank and borrower countries increases. This effect is more pronounced for the geographic distance than for the economic distance.

The increase in credit supply to foreign markets by banks with trading expertise may simply reflect a greater degree of internationalization and a stronger specialization in the lending business of banks with trading expertise compared to banks without trading expertise. Banks with global lending operations may specialize in providing trade credit to exporters from specific markets. For example, the Spanish bank Banco Santander, which we classify as a bank with trading expertise, has a specialization in providing trade credit to Peruvian export firms (see Paravisini, Rappoport, and Schnabl (2015)). De Haas and Van Horen (2012) show that banks generally reduce their credit supply to geographically distant locations. This is consistent with the negative coefficients associated with *log(Distance)* and *Economic Distance* in columns (3) and (4) in Table 2.6.5. However, De Haas and Van Horen (2012) also show that this effect is counteracted if banks operate foreign subsidiaries or foreign branches or have lending experience in a foreign market. Thus, the positive sign of the interactions between *# Trading Memberships* > 2 and *log(Distance)* or *Economic Distance* may simply reflect a geographically more dispersed branch network of banks with trading expertise compared to banks without trading expertise. Unfortunately, our data does not allow us to directly observe the branch network of the banks in our sample. However, LPC DealScan does report some branch information for loans in our sample as an addition to the lender name. For example, for some loans granted by BNP Paribas, DealScan reports "BNP Paribas Singapore Branch" as the lender name. Thus, for each loan in our sample for which we have some indication of the specific branch that granted the loan, we hand-collect the branch country and use the great-circle distance formula to compute the geographic distance between the branch country and the borrower country.[26] While this is a rather imprecise measure of a bank's branch network, it

banks and borrowers from Taiwan (rather counting them as part of China), there is no KOF Globalisation Index published for Taiwan.

[26]Note that in many cases, the bank country, branch country, and borrower country are different. For example, we observe loans granted by BNP Paribas to borrowers in Malaysia or the Philippines via the Singapore Branch of BNP Paribas.

Table 2.6.4: Do Trading Banks Have Geographically More Dispersed Lending Operations Than Non-Trading Banks?

Notes: In this table, we report the average geographic distance between bank country and borrower country for banks with trading expertise and without trading expertise. # Trading Memberships >2 indicates banks with more than two trading memberships at securities exchanges and thus indicates trading expertise. # Trading Memberships ≤ 2 indicates the opposite. Avg. Distance Bank is the mean value of the logarithm of the geographic distance between the bank country and the borrower country. Avg. Distance Branch is the mean value of the logarithm of the geographic distance between the bank branch country and the borrower country. We report significance levels for two-sided t-tests of the mean difference, allowing for unequal sample variance as: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	# Trading Memberships >2	# Trading Memberships ≤ 2	Difference
Avg. Distance Bank	4.813	4.027	0.786^{***}
Avg. Distance Branch	3.271	3.546	-0.274^{***}
Difference	1.541^{***}	0.481^{***}	

may provide us with some general insight regarding the degree of geographic dispersion of a bank's lending business. We report the average distances between the bank country and the borrower country on the one hand and the bank branch country and the borrower country on the other hand in Table 2.6.4. We indeed find that for banks with trading expertise, the average geographic distance between the bank countries and borrower countries is larger than the average geographic distance between the bank branch countries and borrower countries. Additionally, we find that the average geographic distance between the bank branch countries and borrower countries is lower for banks with trading expertise than for banks without trading expertise. Both indicate a greater geographic dispersion of the lending operations and thus a higher degree of internationalization of banks with trading expertise compared to banks without trading expertise. Hence, consistent with the results in De Haas and Van Horen (2012), the increased credit supply of banks with trading expertise to foreign markets seems to be driven by the greater geographic dispersion of the lending operations of these banks compared to banks without trading expertise. Approximately 65.6% of all bank-borrower loan connections in our sample can be described as foreign lending. However, many of these loans are granted within the EEA.[27] However, common regulatory frameworks in many areas and an overall comparatively high

[27]The EEA essentially covers the EU plus Switzerland and Norway.

Table 2.6.5: Is Foreign Lending Affected Differently Than Domestic Lending by Trading Expertise?

Notes: In this table, we present the results regarding the effect of trading expertise on credit supply in foreign lending. The unit of observation is firm cluster-year. Firm clusters are formed based on a firm's country of incorporation, the two-digit SIC code, and a firm's credit rating, estimated based on the median EBIT interest coverage ratios. # Trading Memberships represents the number of a bank's trading memberships at major stock exchanges. # Trading Memberships > 2 equals one if # Trading Memberships is greater than two and zero otherwise. Financial Stress is the value of Financial Stress Indicator, as provided by the US OFR, for a bank's country of incorporation. All regressions include bank-level controls (the logarithm of total assets, return-on-assets, common equity/total assets, cash/total assets, and total loans/total deposits). Foreign lending is a dummy variable that equals one if a bank's and borrower's country of incorporation are not the same. Distance is the physical distance between a bank's and borrower's country of incorporation. Economic distance is the absolute value of the difference in the KOF Globalisation Index of a bank's and borrower's country of incorporation. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: Δlog(Loan Volume)					
	(1)	(2)	(3)	(4)	(5)	(6)
# Trading Memberships > 2	-0.391^{***}	-0.366^{***}	-0.365^{***}	-0.338^{***}	-0.256^{***}	-0.244^{***}
	(0.040)	(0.040)	(0.040)	(0.040)	(0.031)	(0.030)
Foreign lending	-1.501^{***}	-1.491^{***}				
	(0.032)	(0.032)				
log(Distance)			-0.177^{***}	-0.176^{***}		
			(0.004)	(0.004)		
Economic Distance					-0.063^{***}	-0.062^{***}
					(0.002)	(0.002)
Financial Stress		-0.025^{***}		-0.028^{***}		0.002
		(0.009)		(0.009)		(0.009)
(# Trading Memberships > 2)*Foreign Lending	0.420^{***}	0.404^{***}				
	(0.041)	(0.041)				
Foreign Lending*Financial Stress		0.021^{***}				
		(0.006)				
(# Trading Memberships > 2)*Financial Stress		-0.051^{***}		-0.051^{***}		-0.039^{***}
		(0.007)		(0.007)		(0.005)
(# Trading Memberships > 2)*log(Distance)			0.046^{***}	0.044^{***}		
			(0.005)	(0.005)		
log(Distance)*Financial Stress				0.003^{***}		
				(0.001)		
(# Trading Memberships > 2)*Economic Distance					0.010^{***}	0.010^{***}
					(0.002)	(0.002)
Economic Distance*Financial Stress						-0.0004
						(0.001)
(# Trading Memberships > 2)*Foreign Lending*Financial Stress		0.012				
		(0.009)				
(# Trading Memberships > 2)*log(Distance)*Financial Stress				0.002		
				(0.001)		
(# Trading Memberships > 2)*Economic Distance*Financial Stress						0.001
						(0.001)
Observations	268,910	268,910	268,910	268,910	266,317	266,317
Adjusted R^2	0.413	0.409	0.416	0.412	0.386	0.382
Bank Controls	YES	YES	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES	YES	YES
Bank Country-Year FE	YES	NO	YES	NO	YES	NO
Bank Country	NO	YES	NO	YES	NO	YES

degree of economic integration lead to a lower risk of foreign lending for EEA banks to EEA borrowers. For example, it is significantly easier to enforce contracts across borders within the EEA compared to the borders of other countries. Furthermore, there is a comparatively high degree of harmonization of regulations within the EEA. Thus, the EEA might be seen as a single lending market. If we treat the EEA as if it were one country in our definition of foreign lending, the share of bank-borrower loan connections that imply foreign lending is approximately 49.8%. We repeat our analysis of the connection of trading expertise and foreign lending, treating the EEA as a single country, and report the results of this exercise in Table 2.6.6. While the magnitude of some coefficients change, the conclusions remain the same as in our previous analysis.

Table 2.6.6: Is Foreign Lending Affected Differently Than Domestic Lending by Trading Expertise? – EEA

*Notes: In this table, we present the results regarding the effect of trading expertise on credit supply in foreign lending, treating countries in the EEA as one country. The unit of observation is firm cluster-year. Firm clusters are formed based on a firm's country of incorporation, the two-digit SIC code, and a firm's credit rating estimated based on the median EBIT interest coverage ratios. # Trading Memberships represents the number of a bank's trading memberships at major stock exchanges. # Trading Memberships > 2 equals one if # Trading Memberships is greater than two and zero otherwise. Financial Stress is the value of the Financial Stress Indicator, as provided by the U.S. OFR, for a bank's country of incorporation. All regressions include bank-level controls .(the logarithm of total assets, return-on-assets, common equity/total assets, cash/total assets, and total loans/total deposits). Foreign lending is a dummy variable that equals one if a bank's and borrower's countries of incorporation are not the same. Distance is the physical distance between a bank's and borrower's country of incorporation. Economic distance is the absolute value of the difference in the KOF Index of Globalization of a bank's and borrower's country of incorporation. Standard errors are clustered at the bank-firm cluster level. Significance levels: *$p<0.1$; **$p<0.05$; ***$p<0.01$.*

	Dependent Variable: Δlog(Loan Volume)					
	(1)	(2)	(3)	(4)	(5)	(6)
# Trading Memberships > 2	-0.415^{***}	-0.390^{***}	-0.432^{***}	-0.408^{***}	-0.255^{***}	-0.244^{***}
	(0.034)	(0.034)	(0.034)	(0.034)	(0.030)	(0.029)
Foreign Lending	-1.288^{***}	-1.278^{***}				
	(0.031)	(0.031)				
log(Distance)			-0.150^{***}	-0.148^{***}		
			(0.004)	(0.003)		
Economic Distance					-0.062^{***}	-0.061^{***}
					(0.002)	(0.002)
Financial Stress		-0.021^{**}		-0.023^{**}		0.002
		(0.009)		(0.009)		(0.009)
(# Trading Memberships > 2)*Foreign Lending	0.482^{***}	0.462^{***}				
	(0.038)	(0.038)				
Foreign lending*Financial Stress		0.019^{***}				
		(0.006)				
(# Trading Memberships > 2)*Financial Stress		-0.046^{***}		-0.043^{***}		-0.037^{***}
		(0.006)		(0.006)		(0.005)
(# Trading Memberships > 2)*log(Distance)			0.060^{***}	0.057^{***}		
			(0.004)	(0.004)		
log(Distance)*Financial Stress				0.003^{***}		
				(0.001)		
(# Trading Memberships > 2)*Economic Distance					0.010^{***}	0.009^{***}
					(0.002)	(0.002)
Economic Distance*Financial Stress						-0.001
						(0.0005)
(# Trading Memberships > 2)*Foreign Lending*Financial Stress		0.015^{*}				
		(0.007)				
(# Trading Memberships > 2)*log(Distance)*Financial Stress				0.001		
				(0.001)		
(# Trading Memberships > 2)*Economic Distance*Financial Stress						0.001
						(0.001)
Observations	268,910	268,910	268,910	268,910	266,317	266,317
Adjusted R^2	0.407	0.403	0.407	0.404	0.386	0.382
Bank Controls	YES	YES	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES	YES	YES
Bank Country-Year FE	YES	NO	YES	NO	YES	NO
Bank Country	NO	YES	NO	YES	NO	YES

2.7 Conclusion & Policy Implications

Do banks that heavily engage in proprietary trading reduce credit supply in times of crisis more than their peers that are less heavily engaged in proprietary trading? In our analysis, we answer this question using a global dataset containing information on loans granted by 136 leading banks to a wide range of corporate borrowers between 2003 and 2016. We find that banks with greater trading expertise supply less credit than their peers with lower trading expertise during stable times and even less during crisis times. Compared to non-trading banks, trading banks reduce their credit supply by 19% plus an additional 3.25% during crises. Both effects are consistent with theoretical predictions (see Shleifer and Vishny (2010), Diamond and Rajan (2011), Boot and Ratnovski (2016)) and are in line with previous empirical evidence derived from a one-country sample (see Abbassi et al. (2016)). Additionally, we demonstrate that banks engaged in trading also charge higher prices for their loans. Moreover, we show that the global dimension of our analysis is significant. The double effect of trading banks reducing credit supply during periods of crisis and stability can be attributed to US banks. International banks are unique in this regard, as they only reduce their credit supply during crises. From a theoretical point of view, this finding suggests that between US banks and international banks, there are two different channels at work, both leading to lower credit supply. The theoretical model suggested in Boot and Ratnovski (2016) predicts that banks with trading expertise allocate scares funds to scalable short-term securities trading rather than non-scalable long-term relationship lending activities, thus leading to lower credit supply. This channel appears to be at work for US banks but not for international banks. On the other hand, Shleifer and Vishny (2010) and Diamond and Rajan (2011) argue that banks with trading expertise redirect funds from lending to trading during periods of crisis as the returns from investing in distressed assets are higher than returns from lending. This channel appears to be at work both in US and international banks. These differences help in the assessment of differences in the regulatory frameworks regarding proprietary trading in the US and, e.g., the EU, with US regulations being significantly more restrictive than in other countries. Further exploiting our global sample, we also find that while trading banks provide less credit than non-trading banks overall, they tend to provide slightly more credit than non-trading banks abroad. However, during a crisis, trading banks also cut their foreign lending to a greater extent than their non-trading peers. Finally, we show that these spillovers from trading to credit supply have adverse consequences for the real economy as firms have reduced ability to invest in capital and expand their workforce. This last point in particular adds important information to the debate on the new regulations on banks' proprietary trading, as it shows that there are externalities of proprietary trading beyond excessive risk-taking by banks. Therefore, this finding constitutes the first step towards a cost-benefit analysis of regulations that restrict banks in their proprietary trading operations. However, our analysis also shows that real economic impact, while present, is limited. Since our sample consists of borrowers listed on stock exchanges, this suggests that these borrowers have the ability to compensate the reduced bank credit supply by other sources of funding. An extension of our analysis that also includes non-listed borrowers would likely be a fruitful avenue for future research. However, data on non-listed firms is difficult to obtain and is often only available for a limited number of European economies. Overall, our results suggest that the recent regulatory initiatives to separate trading from commercial banking activities, such as lending, are generally well advised, as banks that engage heavily in proprietary trading reduce their credit supply relative to other banks. Moreover, we show that a global perspective matters for the assessment of spillovers from trading to lending.

2.A Chapter Appendix

2.A.1 Control Variables

Table 2.A.1: Trading Expertise and Bank Lending – Control Variables

NNotes: In this table, we present the results for the bank-level controls for the regressions reported in Table 2.5.1. The Capital Ratio is computed as common equity/total assets and the Loans-to-Deposits Ratio as total loans/total deposits. The Liquidity Ratio is computed as cash/total assets. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent Variable: Δlog(Loan Volume)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Total Assets)	0.499^{***}	0.530^{***}	0.531^{***}	0.527^{***}	0.531^{***}	0.480^{***}
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.010)
ROA	2.444	0.776	−0.922	6.546^{***}	−0.968	26.922^{***}
	(2.218)	(2.229)	(2.230)	(1.637)	(2.231)	(1.455)
Liquidity Ratio	0.744^{**}	0.781^{**}	0.417	0.417^{**}	0.415	-0.856^{***}
	(0.302)	(0.304)	(0.297)	(0.205)	(0.297)	(0.190)
Capital Ratio	3.744^{***}	2.967^{***}	3.337^{***}	2.259^{***}	3.349^{***}	-3.419^{***}
	(0.588)	(0.592)	(0.594)	(0.484)	(0.595)	(0.371)
Loans-To-Deposits	0.195^{***}	0.124^{***}	0.150^{***}	0.124^{***}	0.151^{***}	0.194^{***}
	(0.033)	(0.033)	(0.032)	(0.028)	(0.032)	(0.027)
Observations	268,910	268,910	268,910	268,910	268,910	268,910
Adjusted R^2	0.374	0.374	0.374	0.371	0.374	0.192
Firm Cluster-Year FE	YES	YES	YES	YES	YES	NO
Bank Country-Year FE	YES	YES	YES	NO	YES	NO
Firm Cluster FE	NO	NO	NO	NO	NO	YES
Bank Country	NO	NO	NO	YES	NO	YES

Table 2.A.2: Are Trading or Crisis Exposure Affect Capex? - Control Variables

Notes: In this table, we present the results concerning the control variables of the firm cluster level regressions on Capex. The unit of observation is firm cluster-year. The dependent variable is capital expenditure (Capex). The main results can be found in Table 2.5.2. All standard errors are clustered at the firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Capex$_t$				
	(1)	(2)	(3)	(4)	(5)
log(Assets$_{t-1}$)	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Net Debt$_{t-1}$/Assets$_{t-1}$	-0.001	-0.001	-0.0005	-0.0005	-0.0005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Intangible Assets$_{t-1}$/Assets$_{t-1}$	-0.057^{***}	-0.057^{***}	-0.057^{***}	-0.057^{***}	-0.057^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ΔCash$_{t-1}$/Assets$_{t-1}$	-0.051^{***}	-0.051^{***}	-0.051^{***}	-0.051^{***}	-0.051^{***}
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
EBITDA$_{t-1}$/Assets$_{t-1}$	0.612^{***}	0.613^{***}	0.611^{***}	0.609^{***}	0.608^{***}
	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)
Observations	17,768	17,768	17,768	17,768	17,768
Adjusted R^2	0.385	0.386	0.386	0.386	0.386
Country FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Table 2.A.3: Are Trading or Crisis Exposure Affecting Employment Growth? - Control Variables

Notes: In this table, we present the results concerning the control variables of the firm cluster level regressions on employment growth. The unit of observation is firm cluster-year. The dependent variable is employment growth, measured as the year-to-year change in the logarithm of the number of employees. The main results can be found in Table 2.5.3. All standard errors are clustered at the firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Employment Growth$_t$				
	(1)	(2)	(3)	(4)	(5)
log(Assets$_{t-1}$)	0.005^{***}	0.005^{***}	0.005^{***}	0.005^{***}	0.005^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Net Debt$_{t-1}$/Assets$_{t-1}$	-0.018^{**}	-0.018^{**}	-0.018^{**}	-0.018^{**}	-0.018^{**}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Intangible Assets$_{t-1}$/Assets$_{t-1}$	0.110^{***}	0.110^{***}	0.110^{***}	0.109^{***}	0.110^{***}
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
ΔCash$_{t-1}$/Assets$_{t-1}$	0.436^{***}	0.435^{***}	0.435^{***}	0.437^{***}	0.435^{***}
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
EBITDA$_{t-1}$/Assets$_{t-1}$	1.639^{***}	1.641^{***}	1.639^{***}	1.677^{***}	1.678^{***}
	(0.233)	(0.233)	(0.233)	(0.232)	(0.232)
Observations	17,768	17,768	17,768	17,768	17,768
Adjusted R^2	0.060	0.060	0.060	0.061	0.060
Country FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Table 2.A.4: Effect of Trading on Loan Pricing - Control Variables

Notes: In this table, we present the results concerning the bank-level controls for the regressions reported in Table 2.6.1. The Capital Ratio is computed as common equity/total assets and the Loans-to-Deposits Ratio as total loans/total deposits. The Liquidity Ratio is computed as cash/total assets. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Δ*(All-in Spread Drawn)*			
	(1)	(2)	(3)	(4)
log(Total Assets)	-0.728^{***}	-0.714^{***}	-0.725^{***}	-0.715^{***}
	(0.095)	(0.093)	(0.095)	(0.093)
ROA	82.363^{**}	77.198^{**}	84.371^{**}	78.460^{**}
	(39.267)	(38.881)	(39.395)	(38.964)
Liquidity Ratio	-4.796	-5.490	-4.366	-4.806
	(3.868)	(3.856)	(3.914)	(3.926)
Capital Ratio	1.311	1.914	0.816	0.954
	(6.938)	(6.973)	(6.972)	(7.072)
Loans-To-Deposits	-0.324	-0.264	-0.311	-0.264
	(0.347)	(0.354)	(0.348)	(0.354)
Observations	203,947	203,947	203,947	203,947
Adjusted R^2	0.266	0.266	0.266	0.266
Bank Controls	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES
Bank Country	YES	YES	YES	YES

Table 2.A.5: Are U.S. banks different? (Sub-samples) – Control Variables

Notes: In this table, we present the results for the bank-level controls for the regressions reported in Table 2.6.2. The Capital Ratio is computed as common equity/total assets and the Loans-to-Deposits Ratio as total loans/total deposits. The Liquidity Ratio is computed as cash/total assets. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Δlog(Loan Volume)			
	U.S. Banks		non-U.S. Banks	
	(1)	(2)	(3)	(4)
log(Total Assets)	0.832***	0.773***	0.528***	0.534***
	(0.019)	(0.017)	(0.017)	(0.015)
ROA	3.510	23.307***	−14.764***	−0.923
	(3.201)	(2.782)	(2.916)	(1.946)
Liquidity Ratio	1.479**	4.244***	1.310***	0.300
	(0.672)	(0.626)	(0.332)	(0.218)
Capital Ratio	4.155***	−5.713***	0.477	0.894
	(0.988)	(0.613)	(0.775)	(0.583)
Loans-To-Deposits	0.115	0.299***	0.113***	0.084***
	(0.070)	(0.061)	(0.035)	(0.030)
Observations	66,065	66,065	202,845	202,845
Adjusted R^2	0.407	0.272	0.384	0.379
Bank Controls	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES
Bank Country-Year FE	YES	NO	YES	NO
Bank Country	NO	YES	NO	YES

Table 2.A.6: Are U.S. banks different? (Single Equation) – Control Variables

Notes: In this table, we present the results concerning the bank-level controls for the regressions reported in Table 2.6.3. Instead of splitting the sample we introduce a US banks dummy variable and interact it with the # Trading Memberships > 2 dummy, the Financial Stress Index, and all bank-level controls. The upper part of the table shows the estimated coefficients if the US banks dummy equals zero, and the lower part shows the estimated coefficients if the US banks dummy is equal to one. The Capital Ratio is computed as common equity/total assets and the Loans-to-Deposits Ratio as total loans/total deposits. The Liquidity Ratio is computed as cash/total assets. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Δlog(Loan Volume)	
	(1)	(2)
U.S. Banks Dummy = 0		
log(Total Assets)	0.523	0.528^{***}
	(0.015)	(0.015)
ROA	0.573	−0.439
	(1.893)	(1.908)
Liquidity Ratio	0.217	0.330
	(0.211)	(0.213)
Capital Ratio	1.297^{**}	1.433^{**}
	(0.557)	(0.560)
Loans-To-Deposits	0.061^{**}	0.074^{**}
	(0.029)	(0.030)
U.S. Banks Dummy = 1		
log(Total Assets)	0.033	0.026
	(0.023)	(0.023)
ROA	11.299^{***}	12.560^{***}
	(3.144)	(3.611)
Liquidity Ratio	0.464	0.302
	(0.644)	(0.645)
Capital Ratio	−0.915	-1.520^{*}
	(0.772)	(0.790)
Loans-To-Deposits	0.060	0.050
	(0.064)	(0.064)
Observations	268,910	268,910
Adjusted R^2	0.372	0.373
Bank Controls	YES	YES
Firm Cluster-Year FE	YES	YES
Non-U.S. Bank Country Dummies	YES	YES

Table 2.A.7: Is foreign lending affected differently by trading expertise? – Control Variables

Notes: In this table, we present the results concerning the bank-level controls for the regressions reported in Table 2.6.5. The Capital Ratio is computed as common equity/total assets and the Loans-to-Deposits Ratio as total loans/total deposits. The Liquidity Ratio is computed as cash/total assets. Standard errors are clustered at the bank-firm cluster level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

	Dependent variable: Δ*log(Loan Volume)*					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Total Assets)	0.687^{***}	0.675^{***}	0.699^{***}	0.688^{***}	0.582^{***}	0.573^{***}
	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)
ROA	-6.649^{***}	1.080	-7.217^{***}	0.628	-2.497	3.997^{**}
	(2.045)	(1.533)	(2.038)	(1.530)	(2.173)	(1.615)
Liquidity Ratio	2.115^{***}	1.547^{***}	2.249^{***}	1.638^{***}	0.991^{***}	0.744^{***}
	(0.293)	(0.203)	(0.292)	(0.203)	(0.298)	(0.205)
Capital Ratio	4.721^{***}	4.176^{***}	4.806^{***}	4.310^{***}	3.780^{***}	3.050^{***}
	(0.556)	(0.459)	(0.556)	(0.459)	(0.582)	(0.478)
Loans-To-Deposits	0.096^{***}	0.077^{***}	0.083^{***}	0.066^{**}	0.110^{***}	0.093^{***}
	(0.032)	(0.028)	(0.032)	(0.028)	(0.032)	(0.028)
Observations	268,910	268,910	268,910	268,910	266,317	266,317
Adjusted R^2	0.413	0.409	0.416	0.412	0.386	0.382
Bank Controls	YES	YES	YES	YES	YES	YES
Firm Cluster-Year FE	YES	YES	YES	YES	YES	YES
Bank Country-Year FE	YES	NO	YES	NO	YES	NO
Bank Country	NO	YES	NO	YES	NO	YES

2.A.2 Estimating the Fixed-Effects Models on Aggregated Observations

We now discuss the implications of applying our model (2.1) and model (2.2) regressions to firm clusters rather than individual firms. This appendix relies heavily on Veredas and Petkovic (2010). Generally, we are interested in estimating a model in the following form:

$$y_{z,t} = \gamma_z + \beta f_{z,t} + u_{z,t} \tag{2.7}$$

where $z = 1, 2, ..., Z$ indexes individual firms. However, as in our specifications for model (2.1) and model (2.2), we must aggregate individual firms into groups $j = 1, 2, ..., J$ with $J < Z$. Thus, we define an aggregation scheme, such that

$$\tilde{y}_{j,t} = \sum_{z=1}^{Z} M_z^j y_{z,t} \tag{2.8}$$

where $M_z^j = 1$ or 0, such that $\sum_{j=1}^{J} \sum_{z=1}^{Z} M_z^j = J$, i.e., we sum up individuals belonging to group j. We further require $\sum_{z=1}^{Z} M_z^j M_z^{j'} = 0 \; \forall \, a' \neq a$, i.e., that individual firms can only belong to one group. Without loss of generality, we consider a simplified case with only a single independent variable and only individual fixed effects. We focus on a specification equivalent to our model (2.2). All results shown below are easy to apply to our model (2.1) specification.

Applying this aggregation scheme to the regression equation, (2.7) yields

$$\sum_{z=1}^{Z} M_z^j y_{z,t} = \sum_{z=1}^{Z} M_z^j \gamma_z + \sum_{z=1}^{Z} \beta M_z^j f_{z,t} + \sum_{z=1}^{Z} M_z^j u_{z,t} \tag{2.9}$$

$$\tilde{y}_{j,t} = \gamma_j + \beta \tilde{f}_{j,t} + \tilde{u}_{j,t} \tag{2.10}$$

Thus, the slope parameter is not affected by the aggregation, as we assume slopes are constant through individual firms. The group fixed effects γ_j are simply the sum of the individual fixed effects in each group. Note that in terms of our model ((2.1), we have $\tilde{x}_{i,t} = x_{i,t}$, since the control variables are bank-level rather than firm-level variables, and $\sum_{z=1}^{Z} M_z^j = 1$.[28]

[28] Obviously, the same applies to the bank country-year fixed effect.

To understand how the aggregation affects parameter estimation and inference, we write the model in matrix notation.

$$\begin{Bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_\mathbf{Z} \end{Bmatrix} = \begin{Bmatrix} \gamma_1 \mathbf{e}_\mathbf{N} \\ \gamma_2 \mathbf{e}_\mathbf{N} \\ \vdots \\ \gamma_Z \mathbf{e}_\mathbf{N} \end{Bmatrix} + \begin{Bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_\mathbf{Z} \end{Bmatrix} \beta + \begin{Bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_\mathbf{Z} \end{Bmatrix}$$

$$\mathbf{Y} = \mathbf{G} + \mathbf{F}\boldsymbol{\beta} + \mathbf{U} \tag{2.11}$$

where $\mathbf{y_z}$ and $\mathbf{f_z}$ are $(N \times 1)$ vectors containing the observations for individual firm z. γ_z are individual firm fixed effects and $\mathbf{e_N}$ are $(N \times 1)$ vectors of ones. $\mathbf{u_z}$ are $(N \times 1)$ vectors of iid individual firm error terms with $\mathbb{E}(\mathbf{u_z}) = 0$ and $\mathbb{E}(\mathbf{u_z u_z'}) = \sigma_u^2 \mathbf{I_N}$, where $\mathbf{I_N}$ is an identity matrix of size N.

We introduce our aggregation scheme by defining the following matrix:

$$\mathbf{M} = \begin{Bmatrix} M_1^1 & M_2^1 & \cdots & M_Z^1 \\ M_1^2 & M_2^2 & \cdots & M_Z^2 \\ \vdots & \vdots & \ddots & \vdots \\ M_1^J & M_2^J & \cdots & M_Z^J \end{Bmatrix} \tag{2.12}$$

Hence, the aggregation in Equation (2.8) can be written in matrix notation as $(\mathbf{M} \otimes \mathbf{I_N})\mathbf{Y}$. With $\mathbf{A} = (\mathbf{M} \otimes \mathbf{I_N})$, we can write Equation (2.10) as

$$\mathbf{AY} = \mathbf{AG} + \mathbf{AF}\boldsymbol{\beta} + \mathbf{AU} \tag{2.13}$$

Therefore, it follows directly that we have

$$\mathbb{E}(\mathbf{AUU'A'}) = \sigma_u^2(\mathbf{AA'} \otimes \mathbf{I_N}) \tag{2.14}$$

where $\sigma_u^2 = \mathbb{E}(\mathbf{UU'})$. Hence, the aggregation of firms into firm clusters produces heteroscedastic error terms, since the values along the diagonal of $\mathbb{E}(\mathbf{AUU'A'})$ differ.

To estimate the coefficient β we define a standard projection matrix $\mathbf{Q}$ to de-mean observations

$$\mathbf{Q} = \mathbf{I_N} - \frac{1}{N}\mathbf{e_N}\mathbf{e_N'} \tag{2.15}$$

$$\mathbf{\mathcal{Q}} = \mathbf{I_Z} \otimes \mathbf{Q} \tag{2.16}$$

Thus, we have

$$\begin{aligned} \mathbf{\mathcal{Q}AY} &= \mathbf{\mathcal{Q}AG} + \mathbf{\mathcal{Q}AF\boldsymbol{\beta}} + \mathbf{\mathcal{Q}AU} \\ &= \mathbf{\mathcal{Q}AF\boldsymbol{\beta}} + \mathbf{\mathcal{Q}AU} \end{aligned} \tag{2.17}$$

Therefore, it follows directly that the estimated coefficient has the following form:

$$\hat{\beta} = (\mathbf{F'A'\mathcal{Q}AF})^{-1}\mathbf{F'A'\mathcal{Q}AU} \tag{2.18}$$

Since $\mathbb{E}(\mathbf{U}) = 0$, we have $\mathbb{E}(\hat{\beta} - \beta) = 0$, i.e., the estimator is unbiased. However, the estimator is inefficient due to the aggregation and the fixed effects specification, i.e., we have, for the variance of $\hat{\beta}$,

$$\begin{aligned} \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = & \\ &= (\mathbf{F'A'\mathcal{Q}AF})^{-1}\mathbf{F'A'\mathcal{Q}}\mathbb{E}(\mathbf{AUU'A'})\mathbf{\mathcal{Q}AF}(\mathbf{F'A'\mathcal{Q}AF})^{-1} \\ &= \sigma_u^2\left((\mathbf{F'A'\mathcal{Q}AF})^{-1}\mathbf{F'A'\mathcal{Q}}(\mathbf{AA'} \otimes \mathbf{I_N})\mathbf{\mathcal{Q}AF}(\mathbf{F'A'\mathcal{Q}AF})^{-1}\right) \end{aligned} \tag{2.19}$$

2.A.3 Institutional Details

The fact that regulators in many countries have taken action suggests that they believe in the existence of this link. Since the 2007-2009 financial crisis, various regulatory initiatives have been launched to insulate the traditional banking business – such as lending and deposit-taking – from securities trading, including the Volcker Rule in the US, the Banking Reform Act 2013 in the UK, and the Liikanen proposal in the EU.

In the US, the Volcker Rule was introduced in 2010 as part of the Dodd-Frank Act, which prohibits banks from engaging in propriety trading.[29] The Volcker Rule exempts

[29]The rule is named after its author and primary proponent Paul Volcker, who served as chairman of the

certain securities, such as foreign exchange instruments and government securities, hedging, and market-making activities.[30] Since July 2014, banks with trading assets and liabilities worth $50 billion or more have to comply with Volcker Rule regulations, and banks with smaller trading operations are exempt until 2016.[31]

In the UK, the Banking Reform Act 2013, which builds on the Vickers Report, introduced a partial separation of retail banking services from wholesale and investment banking – the so-called "ring-fencing" – to prevent banks from funding securities trading through deposits.[32] The UK government implemented all the necessary legislation in 2015, but the Prudential Regulation Authority has yet to finalize the ring-fencing rules. UK banks are expected to comply with the regulations by 2019 at the latest.[33]

In the EU, the Liikanen proposal[34] suggests two options for reform in concerning securities trading. According to the first option, banks are broken up into separate units, engaging in trading and traditional banking only, if they fail to present to regulators a credible resolution plan, detailing how trading-related activities can be identified and separated during a financial crisis. Moreover, additional non-risk-weighted capital requirements are imposed on banks that engage in securities trading. According to the second option, large and complex banks are broken up by forcing their trading activities into legally separate units. The separate "Trading-houses" may be placed in the same ownership structure, but they must have their own equity and separate funding that cannot come from (government-insured) retail deposit-taking. The implementation of the proposed reforms moves more slowly in the EU. The EU council agreed in 2015 on its position regarding the proposed regulations, which provides the Council President with the necessary mandate to negotiate with the European Parliament on the final version thereof. However, Germany and France independently pushed forward, introducing

Economic Recovery Advisory Board and was former Federal Reserve chairman.

[30]See Duffie (2017).

[31]See, e.g., Krahnen et al. (2017) and Lehmann (2016). Whitehead (2011) provides a comprehensive discussion of the legal concerning the Volcker Rule provisions.

[32]The report is named after Sir John Vickers, then Chair of the UK's Independent Commission on Banking, who authored the report on behalf of the UK parliament.

[33]See Krahnen et al. (2017).

[34]The Liikanen proposal refers to the policy suggestions that were made in the "Report of the European Commission's High-level Expert Group on Bank Structural Reform". The expert group was headed by Erkki Liikanen, governor of the Bank of Finland and member of the ECB council, and it became known as the Liikanen group.

national regulations on banks' trading activities in 2013. Banks in both countries have been required to comply with the regulations since July 2015. Both Germany and France, follow the second option of reform in the Liikanen proposal, allowing banks to continue their securities trading activities but only to exercise them through legally separate entities, which can be placed in the same ownership structure. While largely similar, the German and French regulations differ in the precise definition of the activities from which the non-trading entities are banned.[35] Since the current EU proposal includes a clause that allows national laws to remain in place after the EU regulations become effective, the relationship between EU and national legislation is not clear in this regard.[36]

The above mentioned legislation aims to distinguish proprietary trading, i.e., securities trading on a banks' own account with the intention to profit from the difference between the sales and purchase price, from market-making and hedging. However, Worstall (2013) explains that many transactions that can be framed as proprietary trading share the same characteristics as the maturity transformation (accepting short-term deposits to fund long-term lending) on which traditional banking is built as well as a wide range of hedging activities. Hence, the main challenge faced by regulators tasked with the implementation of the above mentioned rules is to provide a clear and operational definition of the type of securities trading that is to be banned or separated from traditional banking. Furthermore, Duffie (2012) and Duffie (2017) argue that in fact there is no evident distinction between proprietary trading and market-making. Investors rely on the ability of market makers to buy securities or sell them out of their inventory. Most market-making around the world is conducted by bank-affiliated broker-dealers who handle the majority of trading in government, municipality, and corporate bonds as well as over-the-counter derivatives, currencies, commodities, mortgage-related securities, and large blocks of equities. Therefore, enforcing the different regulations regarding propriety trading could lead to a reduction in market-making by banks, potentially leading to losses in market liquidity and eventually a migration of market-making into the less-regulated shadow banking sector.[37] Furthermore, Randal Quarles, current Federal Reserve Vice Chair for Supervision, made a similar point when announcing a review of the Volcker Rule in March 2018, arguing that

[35]See Lehmann (2016).
[36]See Krahnen et al. (2017).
[37]See Duffie (2012) and Duffie (2017).

"We [the Federal Reserve] want banks to be able to engage in market making and provide liquidity to financial markets with less fasting and prayer about their compliance with the Volcker Rule".[38] Consistent with these concerns, Bao, O'Hara, and Zhou (Forthcoming) demonstrate that Volcker Rule-affected broker-dealers reduced market-making activities, leading to lower liquidity in the bond market during periods of stress. Thus, there is a cost-benefit trade-off of regulations regarding securities trading by banks. Our results contribute to the debate by providing evidence for a reduction in credit supply across various jurisdictions as well as negative consequences of banks' securities trading for the real economy.

[38]See Reuters Business News (2018).

3

Is Reported Derivative Use Informative About Risk Taking?[1]

3.1 Introduction

"[T]he true bank balance sheet is itself unobservable. What we actually observe is the accounting balance sheet, which is a quantitative depiction of a bank's economic reality constructed through the application of managerial judgment and discretion to existing accounting rules. Given that regulators and investors make decisions based on what is observable, financial accounting exerts a potentially significant influence on outcomes in the banking sector." – Bushman (2016)

[1]This chapter is based on a working paper co-authored by Jaap Bos (Maastricht University)

Financial derivatives are among the most economically complex financial contracts. By virtue of this complexity, derivatives can be powerful tools for managing and transferring risks in the presence of uncertainty, but they also allow firms to cheaply and easily take speculative risk. Warren Buffet famously emphasized the potential risks of the use of derivatives by describing them as "financial weapons of mass destruction, carrying dangers that, while now latent, are potentially lethal".[2] On the other hand, Robert C. Merton argued in his Nobel lecture that "It's not derivatives that are the problem, it's how they are used" (see Merton (1997)). Determining how derivatives are used is difficult due to their high complexity, but they are especially important in the banking sector, because the derivatives market is traditionally dominated by large banks and due to the systemic importance of banks. Regulators and investors must understand whether banks are using derivatives to increase or manage risks, since derivatives can generate significant losses and even threaten the stability of a bank. For example, AIG suffered $18 billion in derivatives related losses in 2008, and Morgan Stanley and Société Générale lost $9 billion respectively and $7.2 billion in the derivatives market in the same year. Earlier examples of financial institutions that incurred significant losses due to their use of derivatives include Allfirst Bank ($691 million in 2003), Daiwa Bank ($1 billion in 1997), Barings Bank ($1.4 billion in 1995), and Midland Bank ($500 million in 1993).

Since regulators and investors rely on information from financial statements to understand how banks use derivatives, financial accounting rules regarding the reporting of derivatives have a significant influence on the assessment of banks' derivatives use. However, derivatives create significant financial reporting challenges due to their high economic complexity. Consequently, reporting rules for derivatives have become extraordinarily complex and have been referred to by experts as "the poster child of complexity" (see Leone (2007)) and a "labyrinth of processes and documentation" (see Valladares (2014)).

Bank managers can use considerable judgment in navigating this labyrinth, and Kawaller (2004) shows that users of derivatives often do not apply accounting rules relating to

[2] See Warren Buffet's letter of 2003 to shareholders of Berkshire Hathaway, available at `http://www.berkshirehathaway.com/letters/`.

derivatives correctly, or they apply them inconsistently. This potentially drives a wedge between what Bushman (2016) calls the *true balance sheet* and the *accounting balance sheet*, making it difficult for regulators and investors to understand the use of derivatives by banks. Indeed, Chang, Donohoe, and Sougiannis (2016) document that even sophisticated market participants such as sell-side analysts routinely misjudge the implications of firms' derivatives use and show that this misjudgment is driven by the complexity of the reporting of derivatives use rather than their inherent economic complexity.

Therefore, this chapter aims to assess the extent to which the reporting of derivatives use by banks helps regulators and investors to assess a bank's attitude to risk. We show that reported derivatives use is only weakly related to bank-level risk measures and that they tend to associate a larger proportion of hedging derivatives with higher rather than lower bank-level risk. We argue that such an association is likely misleading and an artifact of accounting practices, since a large base of theoretical literature suggests that bank-level risk should decline with increased hedging (see, e.g., Purnanandam (2007), Froot, Scharfstein, and Stein (1993), Diamond (1991), Smith and Stulz (1985), Mayers and Smith (1982)).

Based on corporate risk management theory, we expect to be more likely to observe low levels of risk if a bank uses derivatives mostly for hedging purposes compared to banks that use derivatives mostly for speculative purposes. We suggest an approach to assess banks' derivatives use that is built on this rationale. We regard reported derivatives use as unreliable and instead treat derivatives use as an unobservable, latent variable. In other words, we assume that we can accurately observe the total amounts of derivatives that a bank uses but not *how* these derivatives are used. Employing a latent class regression model then allows us to estimate for each bank the probability that it applies a derivatives strategy focused on hedging or a derivatives strategy focused on speculative trading, conditional on the bank's risk.[3] This approach yields a class of, on average, low-risk hedgers and a class of, on average, high-risk traders. Our latent class regression model indicates that banks are more likely to engage in hedging than in trading and that an assessment of hedging behavior based on reported derivatives use in accordance with hedge accounting rules underestimates the extent to which banks are hedging. We find that while most banks that report a large proportion of hedging derivatives are indeed hedging, there is a large cohort

[3]For an overview of latent class models, see, e.g., Greene (2002), Magidson and Vermunt (2004), Grün and Leisch (2008), Wedel and DeSarbo (1995), Dayton and Macready (1988).

of banks reporting most of their derivative positions as trading while they are in fact engaging in hedging. This result is in line with survey evidence reported in Mulford and Comiskey (2008) and Papa and Peters (2013) suggesting that CFOs prefer to report derivatives as trading rather than hedging instruments even if they are valid economic hedges, because the burden of documentation and reporting complexity are substantially lower for trading derivatives compared to hedging derivatives. This is mainly due to regulators treating "trading purpose" as the default use of any derivative, while banks must produce costly evidence for the presence of an effective hedging relationship to be able to designate a derivative as a hedging instrument (see FinancialCAD & KPMG (2011) and Chranes et al. (2003)). However, banks are incentivized to produce such evidence due to the preferential regulatory and accounting treatment of hedging derivatives that simplifies the smoothing of accounting earnings and allows banks to reduce risk-weights in the computation of equity capital chargers (Gensler (2010), Jorion (2003), Kahane (1977)).

Our results also show that while there is an overall tendency of banks to use derivatives for hedging purposes, an excessively large proportion of derivatives is associated with average risk and, therefore, a tendency for speculative trading. In particular, during the 2007 to 2009 financial crisis, the conditional probabilities of banks using derivatives for hedging declined, but the amounts of derivatives held by banks increased. During the peak of the crisis, we observe the largest number of banks in our sample of US banks changing from hedging to trading or trading to hedging. Banks that changed from hedging to trading during the crisis experienced a sharp decline in their return-on-assets prior to the change, consistent with banks' attempts to increase risk to boost profitability. The banks that changed from trading to hedging during the same time exhibit rather stable return-on-assets despite the financial crisis, consistent with such banks wanting to "lock in" current income.

3.2 Banks' Derivatives Use and Reporting

3.2.1 Have Banks Incentives To Use Derivatives For Hedging?

In a perfect financial market, there would be no incentive to engage in hedging (see Modigliani and Miller (1958)). However, in incomplete financial market, firms are incentivized to engage in hedging, as hedging reduces the costs of external financing and the costs of default (see, e.g., Mayers and Smith (1982),Smith and Stulz (1985), Froot, Scharfstein, and Stein (1993), Dadalt, Gay, and Nam (2002)).[4] In the context of banks, it may be argued that especially for large banks, the incentives to engage in hedging are weaker compared to non-financial firms due to the presence of implicit or explicit government guarantees keeping the costs of external financing low and reducing the costs of default.

However, even if government guarantees for large banks are considered, real costs are incurred by banks' shareholders in the case of a default. These costs mainly arise from a potential loss in the bank's charter value. Loss of charter value can be considered as the loss of stable relationships with clients or business partners, the cost of reorganizations, a loss in reputation, or a loss in organizational sovereignty due to supervisory intervention. Marcus (1984), Keeley (1990), Diamond (1991), and Park (1997) argue that the potential loss of charter value in the case of default incentivizes banks to engage in risk management to reduce the probability of default. Moreover, a lack of liquidity in the event of default may also affect the bank's ability to finance new investments, and it may affect the bank's lending business as new loans cannot be granted (see, e.g., Greenwald and Stiglitz (1993), Shapiro and Titmann (1998), Branch (2001), Hardy (2013)).

Merton (1977) argues that, in the presence of deposit insurance, if the insurance premium paid by banks is insensitive to risk-taking, banks are incentivized to take on more risk as they can exercise the implicit put option. In this case, there would be a weak or no incentive for banks to engage in risk management as long as risk-insensitive government guarantees are available. However, Cyree, Huang, and Lindley (2012) point out that this view implicitly assumes that either default probabilities remain constant or that investors expect a favorable response from the market in light of the higher risk. In particular the latter assumption seems implausible in competitive markets.

Purnanandam (2007) demonstrates that commercial banks' incentives to hedge interest rate risks are an increasing function of the banks' expected costs of default, but the banks' probability to default decreases if it hedges more. Therefore, we expect that low-risk banks are, on average, more likely to be hedgers than high-risk banks.

[4] In the case of a convex tax function, firms can also use derivatives to lower the expected taxes (see Donohoe (2015)).

3.2.2 How Do Banks Report Derivatives Use?

Information on the purpose of banks' derivatives use can be gleaned from supervisory databases. For example, for the US banking system, the data records of banks' quarterly from Form FR-9YC filings with the Federal Reserve contain information on (1) the types of derivatives, (2) gross notional amounts, and (3) a classification into derivatives recorded in the banking book (hedging derivatives) and the trading book (trading derivatives). Generally, the banking book includes all assets that are being held to maturity and without intention to trade, while the trading book consists of all assets held with an intention to trade. Derivatives are classified as trading instruments if they are recorded in the trading book and as hedging instruments if they are recorded in the banking book. This classification is in line with the accounting definition of hedging and trading instruments. Instructions for the filing of Form FR-9YC provided to banks by the Federal Reserve System explicitly mandate the use of this accounting definition of hedging and trading instruments.

According to the Statement of Financial Accounting Standards (SFAS) No. 133, derivatives must be reported at fair value with unrealized gains/losses due to changes in fair value on the income statement.[5] In the context of hedging, this treatment of derivatives can be problematic if the derivatives are used to hedge an exposure that is not subject to fair-value accounting, or if the timing of gains/losses in an exposure does not align with the timing of the gains/losses in derivatives used for hedging. In such a situation, a bank may economically hedge a risk exposure but would not be able to reflect this in its financial statements. Therefore, SFAS No. 133 permits deviation from fair-value accounting to align gains/losses from exposures and derivatives (i.e., so-called *hedge accounting*) if the derivative effectively hedges exposures to (1) changes in the fair value of recognized assets, liabilities, or firm commitments; (2) fluctuations in the cash flows of a recognized asset, liability, or forecasted transaction; or (3) currency risk related to foreign business activities (see Chang, Donohoe, and Sougiannis (2016)). Generally, a hedge is considered effective by the regulator if retrospective (i.e., through back-testing) and prospective (i.e., in terms of the expected effectiveness) hedging positions offset 80% - 125% of the value change in the exposure. The effectiveness of any hedging instrument must be proven anew each quarter

[5] SFAS No. 133 is an accounting standard under US GAAP; however, International Accounting Standard (IAS) No. 39 defines almost identical rules for the accounting of derivatives (see Hughen (2010)).

and, in the case of dynamic hedges, for each adjustment of the position in the derivative (see, e.g., FinancialCAD & KPMG (2011) and Chranes, Koch, and Berkman (2003)). Thus, in order to record a derivative as a hedging instrument in the banking book, a bank must provide evidence that the derivative hedges a recognized exposure and that the hedge is effective. Hedging derivatives according to accounting rules are generally recorded in the banking book, while all other derivatives are generally recorded in the trading book (see, e.g., Jorion (2003)).

3.2.3 Why Would Banks Not Report Derivatives as Hedging Instruments?

SFAS No. 133 is considered the most complex standard ever issued by the US Financial Accounting Standards Board, amounting to approximately 200 pages.[6] Therefore, SFAS No. 133 has been referred to as "the poster child of complexity" (see Leone (2007)) and a "labyrinth of processes and documentation" (see Valladares (2014)). This finding makes it difficult and expensive for banks to designate derivatives as hedging instruments to apply hedge accounting and leads to a lack of transparency in banks' derivatives use.[7] Indeed, Mulford and Comiskey (2008) show in a survey of CFOs across various industries, including banking, that derivatives that are used for hedging are often reported as trading instruments in the firms' financial statements. According to the survey, the reasons for this phenomenon are the substantial costs of documentation and ongoing monitoring of hedges under hedge accounting rules, and the availability of natural hedges that can be highly effective. A 2013 survey by the CFA Institute (see Papa and Peters (2013)) reaches similar conclusions as does the survey undertaken by Mulford and Comiskey (2008). Consistent with this evidence, Ahmed, Kilic, and Lobo (2011) demonstrate, using data on the spreads of banks' bonds and detailed data on derivatives positions concerning the introduction of the SFAS No. 133 hedge accounting rules, that foreign exchange and credit risk-related derivatives that were used by banks for hedging purposes in particular were increasingly reported as trading instruments rather than hedging instruments. Therefore, using publicly

[6]Its international equivalent, IAS No. 39, amounts to approximately 400 pages.

[7]Note that SFAS No. 133 has been amended by other standards in recent years to increase transparency in derivatives use, most notably by SFASs No. 149, No. 155, and No. 161 (see Chang et al. (2016)). However, these additional standards primarily increase the information that is available to outside stakeholders, but do not form the procedure for the application of hedge accounting. Accounting standards relating to the treatment of derivatives are jointly codified in ASC Topic No. 815.

available supervisory data to identify whether banks are using derivatives for hedging or trading purposes may lead to an overestimation of the extent of derivatives trading and consequently an underestimation of the extent of hedging.

Despite the considerable complexity of hedge accounting rules, banks also have regulatory incentives to report derivatives as hedging instruments, since banking book instruments are subject to lower capital charges compared to trading book instruments (see, e.g., Jorion (2003)). Additionally, in the calculation of capital charges for loan or bond portfolios, banks can also take into account whether the risk (e.g., the default, interest rate, or foreign exchange risk) from a loan or bond is hedged. This allows banks to hold less capital for loans and bonds as they can assume a lower exposure to the borrower in the case of default. Banks must add a capital charge for the counter-party risk arising from the derivative itself, though. However, this capital charge is usually comparatively low.[8] Thus, banks can use derivatives to convert loans requiring a high capital charge into ones requiring a low capital charge (see, e.g., Gensler (2010)). Further, Kahane (1977) argues that banks are incentivized to use hedging to minimize the variance of their accounting earnings to reduce the probability of being flagged as risky by supervisors.

3.3 Methodology

We commence this sub-chapter by introducing our supervision-based classification and our model-based classification of banks into hedgers and traders. At the end of the sub-chapter, we then introduce the z-score as our primary measure of banks' risk-taking.

3.3.1 Supervision-Based Classification

Since we obtain information on the level of derivatives that a bank designates as hedging or trading in its financial statements through publicly available supervisory data records, we use the qualifier "supervision-based" to indicate the classifications of banks into hedgers or traders that are derived from the reported derivatives use. The supervision-based classification is based on the most straightforward interpretation of derivatives use reporting in a bank's financial statement. If the total gross notional amount of derivatives

[8]Since the 2007 to 2009 financial crisis, the counter-party risk in derivatives has attracted more attention from regulators.

reported as hedging instruments in bank *i*'s quarter *t* financial statement exceeds that of the derivatives reported as trading instruments, we classify bank *i* as a hedger in quarter *t*. Accordingly, we classify bank *i* as trader in the opposite situation.[9]

While a few banks report exclusively hedging or trading derivatives in their financial statements, most banks report hedging and trading derivatives at the same time. However, if a bank reliably reports each individual derivative – or at least the majority of individual derivatives – according to its actual purpose, the supervision-based classification should yield a reasonable proxy for the tendency of a bank towards hedging or trading. In other words, a bank that is more active in trading than in hedging derivatives likely follows a derivatives strategy that leans towards trading.

3.3.2 Model-based Classification

In order to establish whether the supervision-based classification of hedgers and traders reflects the actual purpose with which banks take derivatives positions, we need a benchmark.

Determining that benchmark is not straightforward, because hedging and trading are activities initiated to achieve exactly the opposite effect on the risk to which a bank is subject: The former is intended to result in a low level of risk, while the latter is intended to increase risk, with the aim of reaping a higher return. Consequently, hedgers and traders have exactly the opposite of the intended effect on bank risk, and pooling hedgers and traders in an empirical benchmark analysis can result in finding no effect. To find out why, consider the following:

$$log(z_{i,t}) = \beta' X_{i,t} + \varepsilon_{i,t} \tag{3.1}$$

where $X_{i,t}$ contains variables that reflect bank characteristics with respect to risk-taking, and $z_{i,t}$ is a measure of the bank's actual risk-taking.[10] Starting with a simple example in which

[9]Ahmed et al. (2011) use a similar approach to classify banks based on their reported derivatives use but classify a bank as a hedger only if it does not report any trading derivatives and as a trader otherwise. Using their classification approach does not change our results qualitatively, but it does lead to a lower number of hedgers.

[10]As is the case in many empirical analyses in finance, in the ideal (but unattainable) scenario, we would want to replace $z_{i,t}$ with *the* measure of *expected* risk taking.

exactly half of the banks are hedgers and the other half are traders, for traders, $\beta' \geq 0$, while for hedgers, $\beta' < 0$. As it stands, whether a bank is *actually* a hedger or trader is not something that we can observe directly; rather, it is unobserved, or latent. In the naïve case that we estimate (3.1) for all banks, still assuming half of them are hedging and the other half trading, we may obtain an estimate of β that is equal to zero, and this tells us nothing about the actual extent to which derivatives positions acted as risk mitigants and, more importantly, about the identity of the hedgers and traders in our sample. In short, we may have fallen victim to the fallacy of composition, assuming that derivatives are not related to a bank's riskiness.

In this chapter, we therefore propose a different, intuitive setup. We start from the premise that hedgers and traders take derivatives positions with the objective of *generating* rather different levels of risk. Indeed, traders aim for a $\beta'_{Trader} \geq 0$, while hedgers are successful if $\beta'_{Hedger} < 0$. In this view, the β we estimate if we naïvely follow Equation (3.1) is a weighted average of β_{Hedger} and β_{Trader}, where the weights reflecting the share of hedgers and traders respectively are unknown. However, we can assume that, in the end, the risk $z_{i,t}$ taken by traders is higher than the risk taken by hedgers. Hence:

$$\mathbb{E}(log(z_{i,t}) \mid X_{i,t}, Hedger) = \beta'_{Hedger} X_{i,t} > \mathbb{E}(log(z_{i,t}) \mid X_{i,t}, Trader) = \beta'_{Trader} X_{i,t} \quad (3.2)$$

If we now also assume that each bank i at a given quarter t is either a hedger or a trader with a certain probability, then for a particular risk value equal to Z, the probability of observing $log(Z) = log(z_{i,t})$ is equal to:

$$\begin{aligned} Prob(log(Z) = log(z_{i,t})) = Prob(log(Z) = & log(z_{i,t}) \mid Hedger_{i,t}) Prob(Hedger_{i,t}) + \\ & Prob(log(Z) = log(z_{i,t}) \mid Trader_{i,t}) Prob(Trader_{i,t}) \end{aligned} \quad (3.3)$$

In that case, we can rewrite Equation (3.1) as a mixture of β'_{Hedger} and β'_{Trader}, and:

$$log(z_{i,t}) = \begin{cases} \beta'_{Hedger} X_{i,t} + \varepsilon_{i,t,Hedger} & \text{if } Prob(Hedger_{it}) \geq 0.5 \\ \beta'_{Trader} X_{i,t} + \varepsilon_{i,t,Trader} & \text{otherwise} \end{cases} \quad (3.4)$$

It is then necessary to determine whether a bank i at in quarter t is more likely to be a hedger ($Prob(Hedger_{it} \geq 0.5$, i.e., $Prob(Trader < 0.5))$, or a trader ($Prob(Hedger_{i,t} < 0.5$, i.e. $Prob(Trader_{i,t} \geq 0.5)$. Realizing that this is a standard latent variable problem, we can estimate the probability of a bank being a hedger (i.e., one minus the probability of being a trader) using a straightforward logit model, where this probability is dependent on the derivatives positions $D_{i,t}$ that a bank uses and a constant, such that:

$$Prob(Trader_{i,t}) = \frac{exp\left(D'_{i,t}\eta\right)}{1 + exp\left(D'_{i,t}\eta\right)} \tag{3.5}$$

$$Prob(Hedger_{i,t}) = 1 - \omega_{Trader}\left(D_{i,t}, \eta\right) \tag{3.6}$$

What results is a system of equations known as a latent class regression model (Magidson and Vermunt (2004), Greene (2002), Wedel and DeSarbo (1995), Dayton and Macready (1988)), consisting of a combination of Equations (3.4) for the estimations of β_{Hedger} and β_{Trader} and a combination of Equations (3.5) and (3.6) for the estimations of $Prob(Trader_{i,t})$ and $Prob(Hedger_{i,t})$. The resulting system is estimated iteratively using a so-called expectation-maximization (EM) algorithm, following Do and Batzoglou (2008), Grün and Leisch (2008), and Bos et al. (2010a). A more detailed discussion of the latent class regression model and the EM algorithm is provided in appendix 3.A.1.

Given the estimated coefficients from this system of equations, the probability of observing $log(Z) = log(z_{i,t})$ in Equation (3.3) is identified. We can apply Bayes' theorem to Equation (3.3) to obtain the conditional probabilities:

$$Prob(Hedger_{i,t} \mid log(Z) = log(z_{i,t})) = p_{i,t,Hedger} \tag{3.7}$$

$$Prob(Trader_{i,t} \mid log(Z) = log(z_{i,t})) = p_{i,t,Trader} \tag{3.8}$$

Finally, we classify bank i in quarter t as a model-based hedger if $p_{i,t,Hedger} > p_{i,t,Trader}$ and as a model-based trader if $p_{i,t,Hedger} < p_{i,t,Trader}$.

3.3.3 Evaluating Goodness-of-Fit

To evaluate the goodness-of-fit of the latent class model, we employ three information criteria: the Entropy Information Criterion (EIC), the Integrated Completed Likelihood Criterion (ICL), and the Bayesian Information Criterion (BIC). EIC is a standardized measure of the entropy of the hedger and trader classes identified by the model.[11] In other words, EIC can be interpreted as a measure of the extent to which the two probability distributions $p_{i,t,Hedger}$ and $p_{i,t,Trader}$ overlap. In an ideal situation, without uncertainty about whether a bank is a hedger or a trader, each of these two probabilities would be either one or zero for any particular observation. However, the more uncertainty there is about a bank being a hedger or a trader, the closer the two probabilities will be to 0.5 for any particular observation. EIC is designed to reflect this rationale. It is bounded between zero and one, whereas values close to one indicate probabilities that are close to zero or one for any particular observation. Values of EIC close to zero indicate probabilities that are close to 0.5 for any particular observation (see Ramaswamy, Desarbo, Reibstein, and William (1993), Pennings and Garcia (2004)). EIC is computed as

$$EIC = 1 - \frac{\sum_{i \in I, t \in T} \sum_{c \in C} - p_{i,t,c} log\left(p_{i,t,c}\right)}{I * T * log(2)} \tag{3.9}$$

where $C = \{Hedger, Trader\}$. I and T denote the number of banks and number of quarters respectively. ICL and BIC share the same interpretations, while ICL equals BIC plus an additional penalty term that takes the entropy of the hedger and trader classes identified by the model into account. For both information criteria, it is ideal to choose a model that minimizes the value of the information criterion. ICL can be computed as

$$ICL = BIC + \sum_{i \in I, t \in T} \sum_{c \in C} - p_{i,t,c} log\left(p_{i,t,c}\right) \tag{3.10}$$

where BIC denotes the usual Bayesian Information Criterion and the second part of ICL equals the mean entropy (see, e.g., Bertoletti, Friel, and Rastelli (2015), Biernacki, Celeux, and Govaert (2000)).

3.3.4 A Measure of Bank Risk-Taking

Faced with an imperfect financial market, banks have an incentive to engage in hedging to reduce their probability of default. Corporate risk management theory suggests that, all else

[11] In statistics and information theory, entropy measures the distance between two probability distributions.

equal, the probability of default of a bank should decrease the more it engages in hedging. In contrast, banks may engage in derivatives trading to increase risk in order to realize higher returns. Therefore, we use the z-score as our primary measure for banks' risk-taking. Note the z-score that is applied in this chapter is not the Altman z-score (see Altman (1968)) commonly applied to measure credit risk in non-financial firms but is not applicable to financial firms such as banks.[12] The z-score that is used in this chapter is a broadly used and well-understood measure of bank risk (see, e.g., Mergaerts and Vennet (2016), Bolton, Mehran, and Shapiro (2015), Köhler (2015), Laeven and Levine (2009), Mercieca, Schaeck, and Wolfe (2007), Boyd, Nicoló, and Jalal (2006)).[13] The z-score is proportional to the inverse of the probability of default, i.e., large z-scores indicate a low probability of default and vice versa and is defined in the following way:

Let $E_{i,t}$ and $\pi_{i,t}$ be the equity and profit of bank i in quarter t respectively. We assume that bank i enters the default state in quarter t if $E_{i,t} < -\pi_{i,t}$, i.e., we define the event of default as a state of the world in which a bank realizes a loss that exceeds its equity capital. The probability of such an event is given by $Prob(ROA_{i,t} < -E_{i,t}/A_{i,t})$, where we simply divide both sides of the inequality by the total assets A and note that $ROA_{i,t} = \pi_{i,t}/A_{i,t}$ is the return-on-assets. Standardizing $ROA_{i,t}$ yields for the probability of default:

$$Prob\left(\frac{(ROA_{i,t} - \overline{ROA})}{\sigma(ROA)} < \underbrace{\frac{(-E_{i,t}/A_{i,t} - \overline{ROA})}{\sigma(ROA)}}_{=-z_{i,t}}\right) = \mathbb{F}\left(-z_{i,t}\right) \tag{3.11}$$

where $\overline{ROA}$ and $\sigma\left(ROA\right)$ are the mean value and standard derivation of the return-on-assets respectively. Figure 3.3.1 visualizes the connection between the negative z-score and the probability of default as defined in Equation (3.11). Following common practice in the banking literature we use the positive values of the z-score. Thus, if profits

[12]The original version of Altman's z-score published in 1968 is only applicable for publicly traded manufacturing firms. Later versions of Altman's z-score are also applicable to private manufacturing firms and non-manufacturing firms (see Altman (2000), Altman (2002)). However, no version of Altman's z-score is recommended for the application to financial firms, due to the particularities of the balance sheet and business model of financial firms compared to non-financial firms.

[13]In the banking literature it is common to refer to the risk measure used in this chapter simply as z-score and to refer to the Altman z-score explicitly as Altman's z-score. Going forward, we apply the same nomenclature.

are distributed normally, we can write the inverse of the probability of default as given in Equation (3.11) as

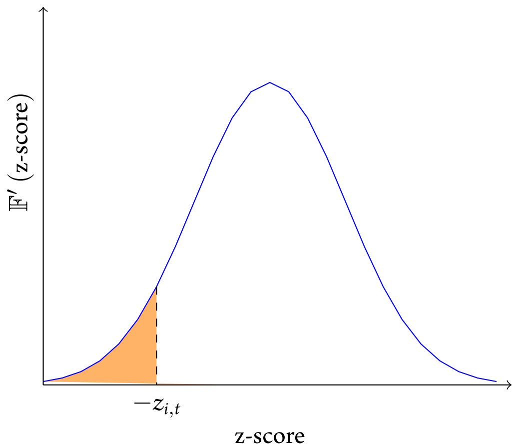
$$z_{i,t} = \frac{\left(\overline{ROA}_{i,t} + E_{i,t}/A_{i,t}\right)}{\sigma\left(ROA\right)_{i,t}} \tag{3.12}$$

In our empirical implementation, we follow Boyd et al. (2006) and compute $\overline{ROA}_{i,t}$ and $\sigma\left(ROA\right)_{i,t}$ as the mean value and standard deviation of the ROA respectively of bank i estimated over a trailing rolling-window of at least eight quarters and up to 12 quarters. We measure $E_{i,t}$ and $A_{i,t}$ as the unweighted book equity and book assets respectively rather than risk-weighted measures, since the risk-weights are directly affected by the derivatives reporting choices of a bank. Profits $\pi_{i,t}$ is net income, including comprehensive income. Since banks can use hedge accounting to reduce volatility in net income but not in comprehensive income, this approach mitigates the effect of hedge accounting on $\pi_{i,t}$ (see Gebhardt, Reichenhardt, and Wittenbrink (2004)). The effect of the reporting choices of banks on the z-score are further mitigated by using the return-on-assets. Since z-scores are highly skewed, we use the logarithm of the z-score, which is distributed normally (see Laeven and Levine (2009)). The logarithm of the z-score is inversely proportional to the odds of default (see, e.g., Lepetit and Strobel (2015)).

The z-score allows for a straightforward interpretation of the effect of hedging on bank risk in this setup. It measures whether a bank is capitalized well enough (average ROA plus equity) to cover profitability fluctuations (standard deviation of ROA). Hedging should primarily affect the standard deviation of ROA and not the bank's capitalization. However, the z-score is a measure of realized default risk, while hedging decisions are based on future expected default risks. Therefore, we implicitly assume that banks' expectations concerning their next quarter default risks are correct.

Figure 3.3.1: Visualization of the z-score

Notes: The figure shows a generic visualization of the connection between the probability of default and the z-score as defined in Equation (3.11). $\mathbb{F}'$ *denotes the density of the probability distribution function* $\mathbb{F}$*, where* $\mathbb{F}$ *describes the probability of default for a specific value of the z-score.*



3.4 Data

Our dataset comprises US bank holding companies. Whenever it does not cause confusion, we use the term "bank" instead of "bank holding company" for simplicity. A bank holding company is a corporation that controls one or more bank subsidiaries but can also have subsidiaries that engage in non-banking activities like insurance, asset management, or securities dealing. In the US, the majority of banks are part of a bank holding company.

All financial statement data is obtained from Form FR Y-9C filings that can be accessed through the Holding Company Database of the Federal Reserve Bank of Chicago. Form FR Y-9C is required to be filed quarterly by all US bank holding companies with consolidated total assets that are larger than $500 million and consolidates all financial statement data of all entities within a bank holding company (see Avraham, Selvaggi, and Vickery (2012)). Form FR Y-9C is required to be filed in accordance with US GAAP and SEC rules. Regarding the recording of derivatives as hedging or trading instruments, the instructions provided by the Federal Reserve System to bank holding companies explicitly mandate that the accounting rules laid out in SFAS No. 133 must be followed. Rampini, Viswanathan, and Vuillemey (2017) show that within bank holding companies, an average of 88.5% of the derivatives exposures are concentrated within the bank's subsidiaries. Thus, derivatives use at the bank holding company level is driven by derivatives use in the bank subsidiaries rather than the non-bank subsidiaries. Using accounting data that is consolidated at the bank holding company level has the advantage that transactions in which both parties belong to the same banking group are netted out.

To construct our sample, we collect all bank holding companies in the Holding Company Database between Q1 1997 and Q4 2015. We drop from this sample all US branches of non-US bank holding companies, bank holding companies that are themselves subsidiaries of another bank holding company, and bank holding companies that report negative book equity. Since Sinkey and Carter (2000) and Cyree, Huang, and Lindley (2012) have demonstrated that banks that use derivatives differ systematically from those that do not use derivatives, we also drop bank holding companies that report zero total gross notional amounts of interest rate, foreign exchange, and credit derivatives for the entire sample period. Restricting the sample in this way, of course, comes at the cost of the reduced external validity of our results. Strictly speaking, we cannot make claims about the full universe of bank holding companies, but only about those that use derivatives in some way. However, the primary interest of this chapter is the reporting of derivatives use by banks and its relationship with bank risk.

We merge the remaining bank holding companies with the CRSP database, using a link-table maintained by Federal Reserve Bank of New York (2017). We drop all non-exchange-listed bank holding companies and those with less than three years of observations in the matched sample. This results in a sample of 454 individual bank holding companies. Detailed variables, definitions, and summary statistics for all variables used in the analysis are provided in the chapter appendix.

3.5 Results

3.5.1 Is Reported Derivatives Use Informative?

Using the latent class regression model described in sub-chapter 3.3, we use our panel dataset to estimate quarterly the probabilities of a bank being either a hedger or a trader conditional on log z-scores. For each quarter, we classify banks as model-based hedgers if their conditional probability of being a hedger exceeds the conditional probability of their being a trader and as model-based traders in the opposite situation. We compare this model-based classification to the supervision-based classification, obtained from the reported use of derivatives. For the supervision-based classification, we classify banks as hedgers if the gross notional amount of derivatives designated as hedges under hedge accounting rules exceeds the gross notional amount of derivatives designated as trading derivatives.

Figure 3.5.1 shows the quarterly averages of the log z-scores and standard deviations of ROA, each with their corresponding 95% confidence bands, for model-based hedgers and traders and supervision-based hedgers and traders. Panel a) of Figure 3.5.1 shows the results for the supervision-based classification. For this classification, there is no statistically significant difference in the log z-scores or ROA volatility for hedgers and traders, suggesting an approximately equal default risk independent of how derivatives are used. This finding suggests that there is no effect of a bank's use of derivatives for either hedging or trading purposes on the banks' probability of default. Given that corporate risk management theory shows that reducing the probability of default is one of the primary motivations to engage in hedging, this conclusion appears rather unlikely.

Panel b) of Figure 3.5.1 shows the results for the model-based classification. The log z-scores of model-based hedgers and traders clearly co-move over time, especially during the 2007 to 2009 financial crisis. The correlation between the log z-scores of the model-based hedgers and traders is 0.803 and statistically significant at all conventional levels. However, the average log z-scores of the model-based hedgers is consistently larger than the average log z-scores of model-based traders, indicating a consistently lower probability of default of the model-based hedgers. This lower probability of default appears to be driven by a consistently lower ROA volatility compared to the model-based traders.

A comparison of panels a) and b) of Figure 3.5.1 shows that the model-based hedger and supervision-based hedger classes are almost identical in terms of average default risk and ROA volatility. However, model-based traders exhibit significantly higher default risk and ROA volatility than supervision-based traders. We now turn to the class-wise regression model. We report the estimated coefficients of the latent class regression model in Equation (3.16) in Table 3.5.1. In the columns labeled "Model-based", we report the estimated class-specific coefficients for the latent class regression model. In the columns labeled "supervision-based", we report the estimated class-specific coefficients using the supervision-based classification. We evaluate the model-based and supervision-based classification in terms of their respective EICs. The EIC for the supervision-based approach is evaluated by using the estimated maximum likelihood coefficients to compute class membership probabilities conditional on the expected z-score comparable to the class membership probabilities of the model-based approach. This can be interpreted as "updating" the observed classification using the expected z-score estimated based on a set of bank characteristics. Updating the supervision-based classification may yield a fairer comparison between supervision-based and model-based classification than the one shown in Figure 3.5.1, as the probability of default may be influenced by a variety of factors.

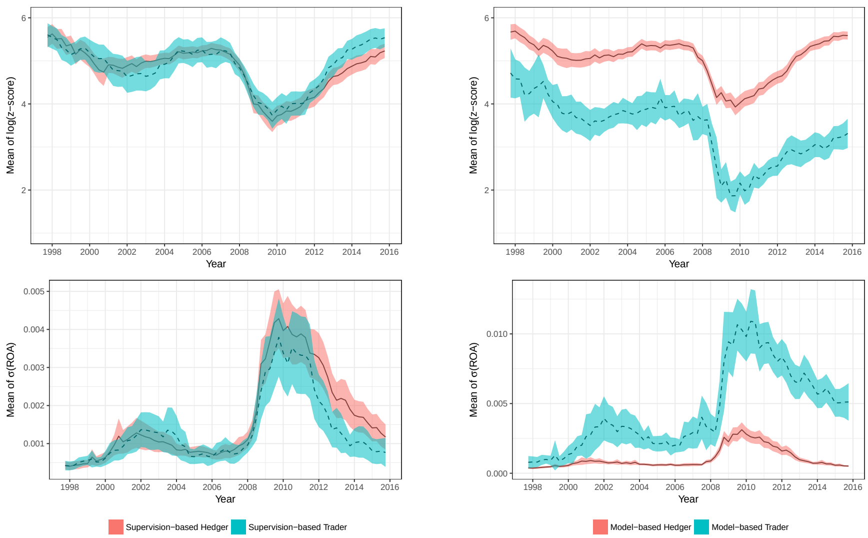
The model-based approach yields an EIC of approximately 0.760, while the supervision-based approach only yields an EIC of approximately 0.118. This implies that the risk updated probabilities of a bank being a supervision-based hedger or a supervision-based trader are close to 0.5-0.5 for a large proportion of the observations. This is an indication that the supervision-based classification does not help in identifying banks

Figure 3.5.1: Can Classifications Distinguish Between Risk-Taking by Hedgers and Traders?

Notes: These figures display the quarterly cross-sectional means of the log z-scores and standard deviations of ROA for hedger and trader banks. Panel a) shows the results for the model-based classification and panel b) shows the results for the supervision-based classification. In both panels, the hedgers are depicted in blue with a dashed line and traders are depicted in red with a solid line. The colored shading represents the 95% confidence bands around the quarterly mean values. The sample consists of a total of 12,593 quarterly observations of 454 bank holding companies from QIV:1997 to QIV:2015. log(z-score) represents the inverse of the odds of default. $\sigma(ROA)$ is the standard deviation of ROA computed over the trailing rolling-windows of eight quarters.

a) Supervision-Based Classification

b) Model-Based Classification



that are engaged in hedging or trading if we assume that engaging in hedging or trading affects bank risk. Using these updated supervision-based probabilities would almost be equivalent to classifying banks into hedgers and traders based on a coin-toss.

On the contrary, the substantially higher EIC of the model-based approach indicates well-separated classes of hedgers and traders with a probability of being a hedger or a trader conditional on the log z-score that is close to either one or zero for most observations. This also implies a rather low classification uncertainty when using these probabilities to classify banks as model-based hedgers or model-based traders.

Lastly, we also compare ICL and BIC for the columns "model-based" and "supervision-based" in Table 3.5.1. Both information criteria are lower for the model-based approach than the supervision-based approach, indicating that the model-based approach describes the observed z-scores better than the supervision-based approach. Table 3.5.1 shows in the columns "model-based" the estimated coefficients of the latent class regression model defined in Equations (3.16) and (3.21). The rows in Table 3.5.1 indicate the bank characteristics $X_{i,t}$ used in Equations (3.16).

Most of the estimated coefficients have the same sign but differ in magnitude for hedgers and traders, suggesting that the z-scores of hedgers and traders co-move as characteristics change, but with different intensities. The confidence intervals for some coefficients overlap for hedgers and traders and in some cases entirely enclose the estimated coefficient of the other class. This finding suggests that not all coefficients are statistically different for hedgers and traders. This is not surprising as not all of the bank characteristics used in the estimation would be expected to be affected by derivatives use. However, in general, the coefficient standard errors are substantially larger for the coefficients of the model-based trader class.[14]

The third of the three columns labeled "model-based" shows the results for the estimation of the parametric function of $Prob(Trader_{i,t})$ as defined in Equation (3.21). The positive signs associated with interest rate derivatives and foreign exchange derivatives indicate that a large proportion of such derivatives are associated with a greater probability of a bank being a trader. For credit derivatives, the opposite applies. Since all derivative

[14]We re-estimate the model in columns "model-based" in Table 3.5.1 but force all coefficients with overlapping confidence intervals to be equal for hedgers and traders. However, specifying the model in this way leads to a lower EIC of 0.38, indicating that model is subject to greater classification uncertainty compared to the current specification. We therefore maintain the current model specification.

Table 3.5.1: Risk Determinants for Hedgers and Traders

Note: In this table, we report the estimated coefficients $\hat{\beta}_{Hedger}$ and $\hat{\beta}_{Hedger}$ and their corresponding standard errors from Equation (3.16). $X_{i,t}$ in Equation (3.16) comprises the variables shown in the left-hand column, an intercept, and a set of year dummies. The coefficients are estimated using the approach summarized at the end of Section 3.3.4. Class Size is based on the count of bank-quarter observations of hedgers and traders over the whole sample period. Class membership in each quarter is assigned based on the maximum posterior membership probability. The reported mean posterior probability is the unconditional pooled mean value of the respective membership posterior probabilities for hedgers and traders. TA refers to total assets, and TL to total loans. The sample consists of 12,593 quarterly observations of 454 bank holding companies (BHCs) from QI:1997 to QIV:2015.

Classification:	**Model-based**			**Supervision-based**	
	$\mathbb{E}(log(Zscore)_{i,t} \mid X_{i,t}, \cdot)$		$\omega_{Trader}\ (D_{i,t}, \cdot)$	$\mathbb{E}(log(Zscore)_{i,t} \mid X_{i,t}, \cdot)$	
	Hedger	Trader	–	Hedger	Trader
Intercept	0.065	-3.350^{***}	-1.860^{***}	-2.053^{***}	0.244^{**}
	(0.085)	(0.210)	(0.099)	(0.237)	(0.105)
IR-Derivatives$_t$/TA$_t$			0.959^{***}		
			(0.133)		
FX-Derivatives$_t$/TA$_t$			1.286^{***}		
			(0.180)		
Credit-Derivatives$_t$/TA$_t$			-22.598^{***}		
			(2.309)		
Size$_t$	0.225^{***}	3.427		1.479^{***}	0.094
	(0.095)	(4.279)		(0.252)	(0.118)
Leverage$_t$	-2.328^{***}	-2.590^{***}		-2.410^{***}	-2.327^{***}
	(0.214)	(0.459)		(0.497)	(0.254)
log(Book-to-Market)$_t$	0.097^{***}	0.146^{***}		-0.070^{*}	0.101
	(0.020)	(0.045)		(0.041)	(0.022)
Intangible Assets$_t$	5.523^{***}	6.612^{***}		10.549^{***}	3.593^{***}
	(0.528)	(1.421)		(1.265)	(0.610)
ROA$_t$	84.006^{***}	3.427		12.455^{**}	90.161^{***}
	(3.304)	(4.279)		(4.895)	(3.951)
Derivatives$_t$	0.003	0.014^{**}		0.046^{***}	-0.018^{***}
	(0.003)	(0.008)		(0.006)	(0.005)
Loans-to-Deposits$_t$	-0.870^{***}	-0.148		-0.851^{***}	-0.609^{***}
	(0.065)	(0.142)		(0.135)	(0.0850)
Total Loans$_t$	0.799^{***}	1.875^{***}		2.574^{***}	0.186
	(0.101)	(0.248)		(0.212)	(0.152)
Demand Deposits$_t$	0.732^{***}	1.020^{***}		1.877^{***}	0.444^{***}
	(0.135)	(0.371)		(0.342)	(0.155)
Non-Performing Loans$_t$	-14.157^{***}	-6.945^{***}		-9.013^{***}	-14.246^{***}
	(0.596)	(1.418)		(1.387)	(0.682)
Loan Loss Reserve$_t$	-23.568^{***}	-24.172^{***}		-33.617^{***}	-23.581^{***}
	(1.608)	(3.098)		(2.969)	(1.980)
Foreign Deposits$_t$	1.464^{***}	1.324^{***}		1.044^{***}	0.091
	(0.200)	(0.299)		(0.404)	(0.197)
Foreign Currency Assets$_t$	-7.717^{***}	0.492		0.267	-9.904^{***}
	(0.696)	(1.219)		(1.285)	(0.903)
EIC	0.760			0.118	
ICL	34039.02			42716.01	
BIC	32059.94			33434.08	

variables are demeaned, the negative intercept implies that a bank with an average proportion of all three types of derivatives is more likely to be a hedger than a trader.

Our main argument is that the supervision-based classification does not help in identifying banks that engage in hedging or trading as their derivatives strategy, because there is no difference in the z-score, i.e., the probability of default, between those banks that report most of their derivatives as hedging and those that report most of their derivatives as trading. However, the z-score itself is only a proxy and is based on certain assumptions. Therefore, to validate or compare model-based and supervision-based classification, we test the differences in the average risk for hedgers and traders using four additional risk measures. Additional to the log z-score, we consider Merton's Distance-to-Default, the Total Stock Return Volatility, Value-at-Risk, and Expected Shortfall.

Merton's Distance-to-Default is also inversely proportional to the probability of default and is equivalent to the z-score. However, the Distance-to-Default is partly based on market inputs rather than accounting variables and is founded in real-option theory. Therefore, it is based on a different set of underlying assumptions compared to the z-score. Value-at-Risk and Expected Shortfall measure the risk of losses in the market value of equity rather than the probability of default. Total Stock Return Volatility measures the riskiness of a bank's equity for investors. Thus, these risk measures capture various aspects of bank risk, are based on market information, and are based on different underlying assumptions.

We test the difference in the mean value of each variable in the hedger and trader class using the Welch t-test, i.e., we test for differences in the mean values, while allowing for different variances in each class. Additionally, we perform a Mann-Whitney test to test for differences in the class-wise distributions. The Mann-Whitney test is non-parametric and thus does not rely on the normality assumption of the t-test. However, the t-test is robust in moderately large samples against violations of the normality assumption (see, e.g., Fagerland (2012)).

A statistically significant Mann-Whitney test, indicating different distributions, can result even if the mean and median values of the considered variable are identical in two classes, but the variances in the classes differ (see Fagerland (2012)). This situation makes it difficult to interpret a statistically significant Mann-Whitney test, as it is not clear exactly what statistical significance implies. Indeed, the Mann-Whitney test statistic is valid under a number of different null hypotheses (see, e.g., Fay and Proschan (2010)). This makes it necessary to make a specific assumption regarding the null hypothesis that is tested. In this chapter, we focus on the location shift hypothesis, i.e., we assume that under the null hypothesis, the distributions in the hedger and trader class of the respective variable are the same against the alternative that one's class distribution is shifted in location (see Perspective 6 in Fay and Proschan (2010)). Therefore, the Welch t-test complements the Mann-Whitney test, as a significant t-test implies provides evidence against the hypothesis of equal means. If both tests are significant, then it is likely that the respective risk measure in each class originates from two different distributions.

Panel a) of Table 3.5.2 shows the results for the model-based classification, and panel b) shows the results for the supervision-based classification. For the model-based classification, the Welch t-tests and Mann-Whitney tests are significant for all risk measures. This finding suggests that the distributions of the respective risk measures are shifted between model-based hedgers and traders. The results for all risk measures are consistent, indicating a higher average risk of model-based traders compared to hedgers.

The results for the supervision-based classification are shown in panel b) of Table 3.5.2. The t-test for the difference in Merton's Distance-to-Default is not significant, while the difference in the Mann-Whitney test is significant. This finding suggests that the distributions are likely not shifted; rather, they may have different variances. We obtain the same results for the Expected Shortfall. For Total Stock Return Volatility and Value-at-Risk, the differences between hedger and trader classes are significant, but the absolute differences are rather small. Further, risk appears lower in the supervision-based trader class compared to the supervision-based hedger class. Since the supervision-based hedger class consists of banks that report most of their derivatives as hedging instruments under current reporting rules, this finding may suggest uncertainty by market participants about the reported derivatives use.

Table 3.5.2: Market-Based Risk Measures

Note: This table shows five risk measures: the logarithm of the z-score, Merton's Distance-to-Default, Total Stock Return Volatility, Value-at-Risk, and Expected Shortfall. The reported mean and median values are pooled over all quarters. In the t-Test column, we report the p-value of a two-sided t-test of the null hypothesis of the equal means of the respective risk measures in the hedger and trader classes, allowing for unequal variance in the two classes. In the Mann-Whitney column, we report the p-value of a Mann-Whitney-U test of the null hypothesis that the respective risk measures have the same distribution in each class against the alternative hypothesis of shifted distributions. Merton's Distance-to-Default is estimated using the approach suggested in Shumway (2001). Total Stock Return Volatility is the standard deviation of the daily log total returns of individual bank stocks within a given quarter. Value-at-Risk and Expected Shortfall are both computed at the 1% level using the historical distribution of the daily log total returns of individual bank stocks within a given quarter.

Panel A: Model-based Classification

	Hedger		**Trader**		**p-Value**	
	Mean	Median	Mean	Median	t-Test	Mann-Whitney
Distance-to-Default	3.332	3.125	2.976	2.521	0.000	0.000
Total Stock Volatility	0.023	0.018	0.027	0.020	0.000	0.000
Value-at-Risk	−0.052	−0.040	−0.061	−0.045	0.000	0.000
Expected Shortfall	−0.062	−0.046	−0.073	−0.052	0.000	0.000

Panel B: Supervision-based Classification

	Hedger		**Trader**		**p-Value**	
	Mean	Median	Mean	Median	t-Test	Mann-Whitney
Distance-to-Default	3.273	3.034	3.320	3.060	0.225	0.019
Total Stock Volatility	0.024	0.036	0.023	0.036	0.030	0.010
Value-at-Risk	−0.054	−0.040	−0.052	−0.040	0.046	0.025
Expected Shortfall	−0.064	−0.047	−0.062	−0.046	0.164	0.064

In Table 3.5.3, we show within the model-based hedgers class and model-based traders class the share of observations that are supervision-based hedgers or traders. We find that

Table 3.5.3: How Are the Reported Classes Distributed Over the Predicted Classes?

Notes: This table shows the proportion of the bank-quarter observations of model-based hedgers and model-based traders that are classified as supervision-based hedgers or supervision-based traders respectively. Thus, 11.33% in the first row and second column means that 11.33% of the observations that are classified as supervision-based hedgers are model-based traders.

	Model-Based Hedger	Model-Based Trader
Supervision-Based Hedger	**88.67%**	11.33%
Supervision-Based Trader	82.02%	**17.98%**

88.67% of observations that are classified as model-based hedgers by our latent class model reported most of their derivatives as hedging instruments and are therefore also classified as supervision-based hedgers. Thus, consistent with the idea that banks that follow a derivatives strategy that is focused on hedging rather than trading should have a lower probability of default, these banks appear to indeed be engaging in hedging. Only for 11.33% of the observations do banks report most of their derivatives as hedging but are identified by our latent class regression model as traders.

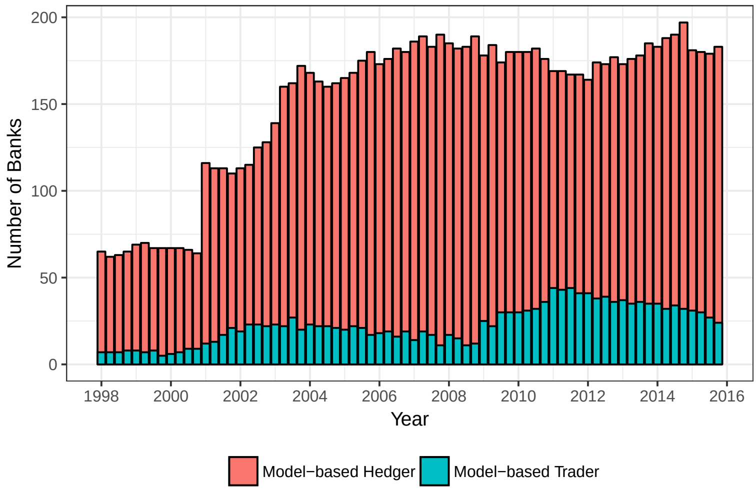
However, 82.02% of observations in the supervision-based trader class are identified by our model as hedgers, i.e., even though the banks report most of their derivatives as trading instruments, these observations of the banks have risk characteristics consistent with a derivatives strategy that is focused on hedging. Only 17.98% of the observations in the supervision-based trader class appear to be consistent with a derivatives strategy that is focused on hedging. This suggests that the supervision-based classification likely underestimates the extent to which banks are hedging. Our results are consistent with the survey evidence provided in Mulford and Comiskey (2008) and Papa and Peters (2013) in this regard. This finding suggests that the costs-associated requirements for banks to be able to designate derivatives as hedges and the rather narrow definition of hedging applied by regulators lead banks to leave derivatives that are valid economic hedges designated as trading instruments. The statistics in Table 3.5.3 are the percentages of bank-quarter observations of the supervision-based classes. Figure 3.5.2 shows instead the sizes of the model-based and supervision-based hedger and trader classes in terms of the number of individual banks in each class. Consistent with Table 3.5.3, we see in Figure 3.5.2 that the model-based hedger and the supervision-based hedger class are of similar size, while the supervision-based trader class is larger than the model-based trader class. The figure also shows an overall trend towards increased derivatives use by banks as the sample size grows over time. However, the conclusion regarding this increased scope of derivatives use differs for supervision-based and model-based classification. For model-based classification, most new entrants into the sample, i.e., banks that have recently chosen to use derivatives are classified as hedgers, suggesting that these banks are characterized by relatively low risk and a derivatives strategy that is focused on hedging. However, for supervision-based classification, we observe that most new entrants are classified as supervision-based traders, as they report most of their derivatives as trading instruments. This finding may be indicative of the high complexity of derivatives reporting rules and documentation requirements for designating derivatives as hedging instruments. In particular, new derivatives users may lack the experience in navigating the "labyrinth of processes and documentation" (see Valladares (2014)) associated with hedge accounting and therefore simply report most of their derivatives as trading instruments. Figure 3.5.3 shows for each quarter the number of individual banks that are classified as model-based hedgers and at the same time as supervision-based traders and vice versa. Since the early 2000s, the number of banks for which the model-based classification and supervision-based classification differ has steadily increased. Generally, there are more banks that are classified as model-based hedgers and supervision-based traders than the other way around.

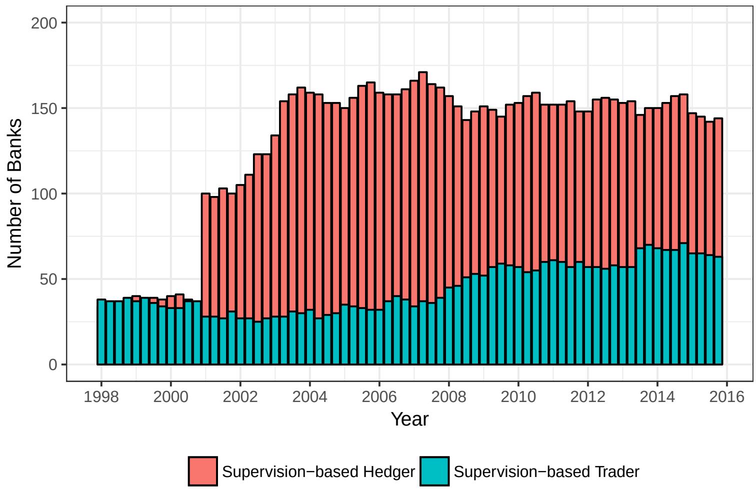
Banks were required to apply the current accounting rules of SFAS No. 133 for designating derivatives as hedging instruments in supervisory reporting by the end of 2001. Ahmed et al. (2011) reference various consultation letters submitted by banking interest groups and individual banks to the Federal Reserve System voicing concerns that the rules of SFAS No. 133 would make it difficult for banks to properly report their hedging activities in their financial statements. Indeed, we find that the number of banks that report most of their derivatives as hedging but are nevertheless classified as model-based hedgers has consistently increased since 2002. This finding is consistent with the assertion that current hedge accounting rules impose a burden that is too large and therefore lead banks to under-report their hedging activities.

The number of banks that are classified as supervision-based hedgers and as

Figure 3.5.2: How Many Banks are Hedgers or Traders?

Notes: These figures show the number of banks that are classified as supervision-based hedgers and supervision-based traders per quarter (upper panel) and model-based hedgers and model-based traders (lower panel).

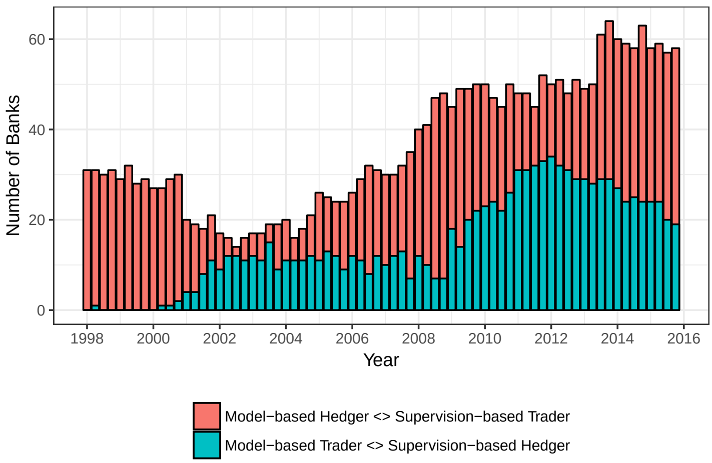




model-based traders simultaneously increased between 2000 and 2002, but then remains rather stable until 2009. Since 2012, the number has been declining. The sharp increase between 2009 and 2012 was likely driven by the financial crisis, which created strong incentives for banks to create the appearance of safety but on the other hand to also engage in trading as opportunities for profitable but risky trading generally increased during this time of crisis (see, e.g., Abbassi et al. (2016), Diamond and Rajan (2011), Arping (2013)). We further investigate the idea that the large number of banks that are supervision-based

Figure 3.5.3: Mis-Classifications

Notes: This figure shows the number of banks per quarter that are model-based hedgers but supervision-based traders in the same quarter and vice-versa.



traders and at the same time are classified as model-based hedgers is driven by banks under-reporting their hedging activities by comparing their derivatives positions. In particular, we compare the gross notional amounts of derivatives divided by the banks' total assets for different derivative types and different reported derivative uses in each class. We distinguish five overlapping types of derivatives (Exchange Traded, Over-the-Counter, Interest Rate, Foreign Exchange, and Credit) and two uses of derivatives (hedging or trading). In Table 3.5.4, we show the averages for supervision-based hedgers and supervision-based traders. As indicated by the values > 1, supervision-based traders appear to hold an excessively large proportion of derivatives, with the average gross notational level of total derivatives exceeding the total assets.[15] However, there is considerable variation among the total gross amount of derivatives held by supervision-based traders. Generally, the largest proportions of derivatives as shares of total assets are Over-the-Counter and Interest Rate-related derivatives. Banks' derivatives positions are dominated by interest rate derivatives, followed by Foreign Exchange-related derivatives. Credit-related derivatives account for only a small fraction of the derivatives positions. By construction supervision-based hedgers possess a larger proportion of hedging derivatives as a share of the total assets, while supervision-based traders possess a larger proportion of trading derivatives as a share of the total assets. Almost all Foreign Exchange- and Credit-related derivatives are held by supervision-based traders, with only incremental levels of these derivatives being designated as hedges. Table 3.5.5 follows the same logic as Table 3.5.4 but also shows the averages of the gross notional amounts of derivatives divided by the total assets for model-based hedgers and traders for the same derivative types and uses as before. The overall pattern is the same here as for the supervision-based classification. Model-based traders possess a significantly larger proportion of derivatives than model-based hedgers, as the gross notional amounts of derivatives exceed the total assets. The derivatives positions are again dominated by Over-the-Counter, Interest Rate-related derivatives, followed by Foreign Exchange-related derivatives, and finally Credit-related derivatives. However, model-based hedgers possess, on average, a significantly larger proportion of total derivatives as a share of the total assets that are designated as trading instruments than those that are designated as hedging instruments. We observe the same pattern across Interest Rate-, Foreign Exchange-, and Credit-related derivatives. This again is consistent with banks not designating derivatives as hedging instruments even though they are valid economic hedges due to the large bureaucratic burden imposed by hedge accounting rules for designating derivatives as hedges.

[15]Note that this is only possible as we consider gross notional amounts, since we do not have sufficient data to apply netting. Nevertheless, the gross amounts can be seen as a measure of derivatives activity but not as the actual net value of the derivatives.

Table 3.5.4: Derivative Positions of Supervision-Based Hedgers and Traders

Notes: This table shows the mean gross notional amounts of derivatives by derivative type and use for supervision-based hedgers and traders, the corresponding standard deviations, and the differences in mean values along with the significance level for a Welch t-test. In columns (1) and (2), we report the mean positions of different types of derivatives, and the types partly overlap. Column (3) shows the difference between the hedger and trader class means. The statistical significance levels for the mean differences are tested using a two-sided Welch t-test. The significance levels are reported as $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. Standard deviations are reported in parentheses. All values are the gross notional amounts of derivatives divided by the total assets.

Supervision-based	Hedger Mean (1)	Trader Mean (2)	t–Test (2)-(1) (3)
Total Derivatives	0.089	2.348	2.26***
	(0.262)	(6.247)	
Exchange Traded Derivatives	0.009	0.246	0.238***
	(0.120)	(0.779)	
Over-the-Counter Derivatives	0.082	2.224	2.142***
	(0.184)	(6.068)	
Trading Derivatives	0.009	2.228	2.219***
	(0.056)	(6.036)	
Hedging Derivatives	0.073	0.069	−0.004
	(0.149)	(0.126)	
Interest Rate Derivatives	0.085	1.843	1.757***
	(0.258)	(5.147)	
Interest Rate Derivatives Tr.	0.008	1.768	1.76***
	(0.054)	(5.087)	
Interest Rate Derivatives Hd.	0.071	0.063	−0.008***
	(0.145)	(0.121)	
Foreign Exchange Derivatives	0.003	0.409	0.406***
	(0.027)	(1.011)	
Foreign Exchange Derivatives Tr.	0.001	0.403	0.402***
	(0.005)	(1.002)	
Foreign Exchange Derivatives Hd.	0.001	0.005	0.004***
	(0.006)	(0.011)	
Credit Derivatives	0.000	0.086	0.086***
	(0.002)	(0.377)	
Credit Derivatives Trading	0.000	0.043	0.043***
	(0.001)	(0.189)	
Credit Derivatives Hedging	0.000	0.001	0.001***
	(0.0004)	(0.003)	

Table 3.5.5: Derivative Positions of Model-Based Hedgers and Traders

Notes: This table shows the mean gross notional amounts of derivatives by derivative type and use for model-based hedgers and traders, the corresponding standard deviations, and the differences in the mean values along with the significance levels for a Welch t-test. In columns (1) and (2), we report the mean positions of the different types of derivatives, and the types partly overlap. Column (3) shows the differences between the hedger and trader class mean values. The statistical significance for the mean differences are tested using a two-sided Welch t-test. The significance levels are reported as $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. Standard deviations are reported in parentheses. All values are the gross notional amounts of derivatives divided by the total assets.

Model-based	Hedger Mean (1)	Trader Mean (2)	t–Test (2)-(1) (3)
Total Derivatives	0.479	1.908	1.428***
	(3.088)	(4.475)	
Exchange Traded Derivatives	0.029	0.335	0.306***
	(0.191)	(1.021)	
Over-the-Counter Derivatives	0.455	1.792	1.337***
	(2.912)	(4.680)	
Trading Derivatives	0.415	1.700	1.285***
	(2.957)	(4.438)	
Hedging Derivatives	0.063	0.129	0.066***
	(0.111)	(0.267)	
Interest Rate Derivatives	0.397	1.453	1.057***
	(2.523)	(3.750)	
Interest Rate Derivatives Tr.	0.333	1.293	0.96***
	(2.491)	(3.709)	
Interest Rate Derivatives Hd.	0.060	0.125	0.065***
	(0.106)	(0.266)	
Foreign Exchange Derivatives	0.06	0.413	0.353***
	(0.419)	(0.997)	
Foreign Exchange Derivatives Tr.	0.057	0.408	0.351***
	(0.414)	(0.989)	
Foreign Exchange Derivatives Hd.	0.002	0.004	0.002***
	(0.008)	(0.010)	
Credit Derivatives	0.025	0.006	−0.018***
	(0.209)	(0.043)	
Credit Derivatives Trading	0.012	0.003	−0.009***
	(0.104)	(0.0216)	
Credit Derivatives Hedging	0.0002	0.0002	0.00001
	(0.002)	(0.001)	

3.5.2 Are Banks Changing Their Derivatives Strategies Over Time?

Since our empirical model allows banks to move across classes over time, banks are not necessarily classified as model-based hedgers or model-based traders over the entire sample period. Furthermore, in supervision-based classification, banks are re-classified in each quarter as gross notational amounts of derivatives designated as hedging or trading do not remain constant over time.

In model-based classification, the conditional probabilities of a bank being a hedger or trader change over time. Panel a) of Figure 3.5.4 shows the quarterly averages of the gross notational amounts of derivatives divided by the total assets for Interest Rate-related derivatives, Foreign Exchange-related derivatives, and Credit-related derivatives over all classifications. Panel b) of Figure 3.5.4 depicts the average conditional probability of a bank being a model-based hedger and the corresponding 95% confidence band. Between QIV:1997 and QIV:2000, the correlation between the average conditional probability of a bank being a model-based hedger and the average gross notational amounts of Interest Rate-related derivatives is 0.32 but statistically insignificant (t-statistic: 1.12). For Foreign Exchange-related derivatives, the correlation with the average conditional probability of a bank being a model-based hedger is 0.15 (t-statistic: 0.50).[16]

However, between QI:2001 and QIV:2015, the same correlation is -0.82 and highly statistically significant (t-statistic: -10.95). For Foreign Exchange- and Credit-related derivatives, the correlations with the average conditional probability of a bank being a model-based hedger are -0.57 (t-statistic: -5.33) and -0.56 (t-statistic: -5.18) respectively.[17] The decline in the conditional probability of a bank being a hedger appears to lead the increase in derivatives use in the post-crisis period. The decline in the average conditional probability of a bank being a model-based hedger does not necessarily imply that more banks are actually being classified as traders, since the average probability remains firmly above 0.5. Therefore, we investigate the actual movement of banks across classes over time. Figure 3.5.5 shows the number of banks moving from being hedgers to traders or vice versa in each quarter for the model-based classification in the upper panel and for the supervision-based classification in the lower panel. The total number of banks in

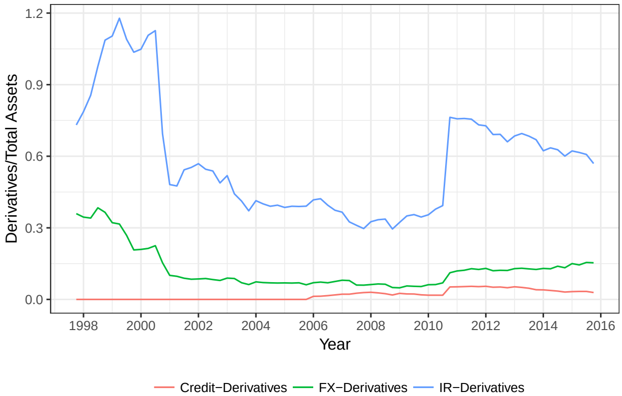
[16]The gross notational amounts of credit derivatives are zero for all banks in the sample during this particular time period.

[17]The conditional probabilities of a bank being a trader exhibit the opposite pattern.

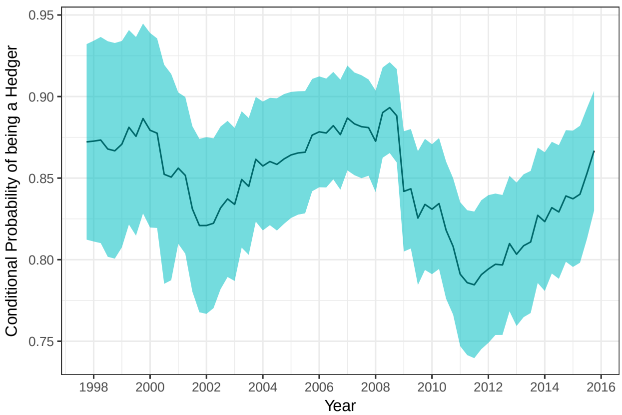
Figure 3.5.4: More Derivatives, Less Hedging?

Notes: Panel a) shows the quarterly averages of the gross notational amounts of Interest Rate-related derivatives (IR-Derivatives), Foreign Exchange-related derivatives (FX-Derivatives), and Credit-related derivatives, each divided by the total assets. Panel b) shows the quarterly average of the conditional probability of a bank's classification as a hedger and the corresponding 95% confidence band.

a) Gross Notational Amounts of Derivatives



b) Conditional Probability of a Bank Being a Hedger

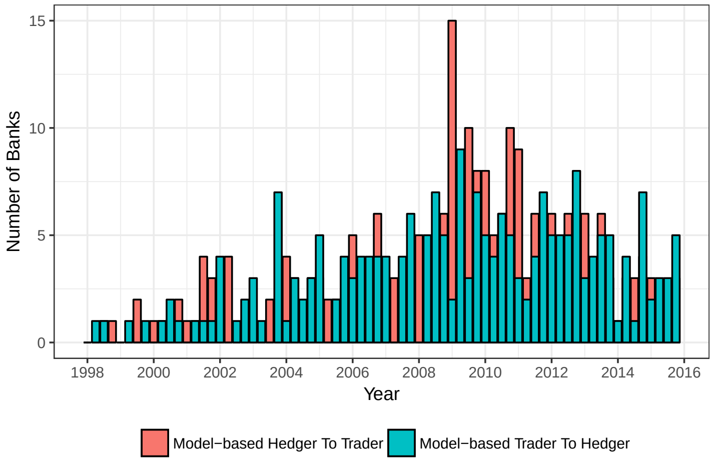


each class for model-based and supervision-based classification are shown in Figure 3.5.2. Overall, the assignments of banks into classes of model-based hedgers and traders and supervision-based hedgers and traders is rather stable over time, with only a comparatively small number of banks moving from one class to the other in each quarter. The average number of moves across classes per bank in the model-based classification is 2.057, with a median of moves across classes per bank of one. This finding shows that our model-based classification indicates a bank-level derivatives strategy towards hedging or trading.

In supervision-based classification, we observe an almost identical stability in assignments of banks into supervision-based hedgers and traders. The average number of moves across classes per bank in supervision-based classification is 1.69, with a median of moves across classes per bank of one. Thus, the reporting choices of banks also appear to be rather stable over time. Therefore, supervision-based classification may be interpreted as a reporting strategy rather than a derivatives strategy. Indeed, in model-based classification, moves across classes appear to cluster in time around the time of the 2007 to 2009 financial crisis, while there is no such pattern in supervision-based classification. Most moves from the model-based hedger class to the model-based trader class or vice versa occur during the financial crisis or the months immediately leading up to it from the first quarter of 2007 to the fourth quarter of 2009. Figure 3.5.6 shows for banks moving from one class to the other the quarterly average annualized ROA of the quarter prior to the move. For banks moving from the model-based hedger class to the model-based trader class, the average ROA declines during the crisis and eventually turns negative. This pattern is consistent with banks changing from a rather safe derivatives strategy that is focused on hedging towards a riskier derivatives strategy that is focused on trading when pressure on profitability increases. On the other hand, the average ROA of banks moving from the model-based trader class to the model-based hedger class remain mostly stable, even during the crisis. This is consistent with banks abandoning rather risky derivatives strategies that are focused on trading as the macroeconomic environment deteriorates and "locking-in" current profitability in the form of a change towards a derivatives strategy that is focused on hedging. While this interpretation is, of course, rather circumstantial and movements across classes may be influenced by other factors as well, it yields a sensible rationale.

Figure 3.5.5: Do Hedgers Become Traders (and Vice Versa)?

Notes: Panel a) shows the number of banks that move from being model-based hedgers in quarter $t-1$ to model-based traders in quarter t and vice versa (upper half) and the number of banks classified as model-based hedgers and model-based traders per quarter (lower half). Panel b) shows the number of banks that move from being supervision-based hedgers in quarter $t-1$ to supervision-based traders in quarter t and vice versa (upper half) and the number of banks classified as supervision-based hedgers and supervision-based traders per quarter (lower half).



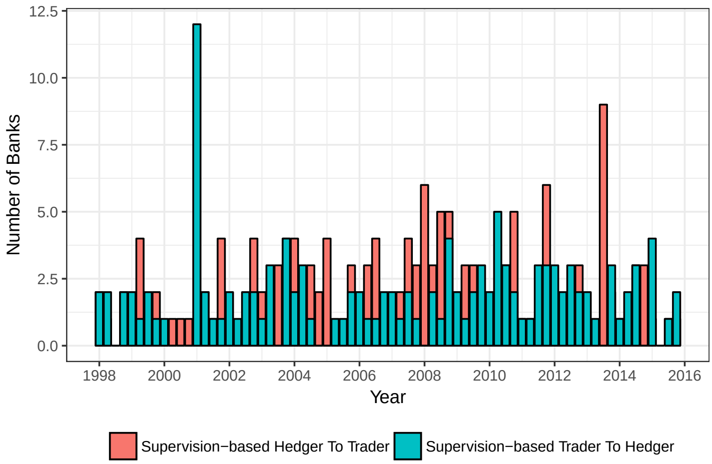


Figure 3.5.6: Does Changing Derivatives Strategy Affect Profitability?

Notes: This figure shows the annualized average of the quarter's ROA for banks that moved from one model-based class to the other during the period of crisis from 2007 to 2009. Banks move from one class in quarter t to the other class in quarter $t+1$, and the figure shows the average ROA in quarter t.

3.6 Conclusion

Bank managers are required to exercise a considerable amount of judgment when applying hedge accounting rules to the designation of derivatives as either trading or hedging instruments. If a bank decides to designate a derivative as a hedging instrument, it must provide extensive evidence regarding the effectiveness of the hedge relationship and must comply with extensive documentation requirements. In surveys, banks claim that this leads to under-reporting the extent of hedging activities, as the bureaucratic burden of designating a derivative as hedging often outweighs the benefit. Together with the overall extraordinarily high complexity of derivatives use reporting, this phenomenon results in a situation in which it is unclear whether reported derivatives use helps regulators and investors in the assessment of banks' derivatives use and whether such use is related to banks' attitudes to risk.

We demonstrate that reported derivatives use under current hedge accounting rules is only weakly related to bank risk. Therefore, reported derivatives use does not provide investors with an indication of whether a bank applies a derivatives strategy that is focused on either hedging various risks or one that is focused on speculative trading. While reported derivatives use tends to underestimate the extent of hedging activities in the banking sectors, it appears that it also underestimates the risk of those banks reporting most of their derivatives as trading instruments.

In this chapter, we employ a latent class regression model to estimate, for each bank in our sample, the probability that the bank is following a derivatives strategy that is focused on hedging conditional on the bank's default risk. This model-based classification of banks into hedgers and traders follows from the results of applying corporate risk management theory, which show that a bank's probability of default should decline as it hedges more. Based on the latent class regression model, we identify a class of hedger banks that is characterized by a comparatively low default risk and a small proportion of derivatives and a class of trader banks that is characterized by a comparatively high default risk and a large proportion of derivatives.

A comparison of a classification of banks based on reported derivatives and a classification based on our latent class regression model indicates that banks do appear to under-report the extent of their hedging activities. While most banks that report most of their derivatives as hedging instruments are indeed hedging, there is a large group of banks that reports most of their derivatives as trading instruments, but they are in fact hedging.

Creating reporting rules for derivatives such that there is an incentive for banks to under-report hedging may be reasonable from a regulatory perspective as it results in a "worst-case" picture of banks' derivatives use. However, the US Financial Accounting Standards Board's Statement of Principles published in 1999 declares that "the objective of financial statements is to provide information [...] useful for assessing stewardship [...] and for making economic decisions." Our results suggest that current reporting rules for derivatives use are inconsistent with this principle.

3.A Chapter Appendix

3.A.1 Latent Class Regression Modeling

In line with the earlier theoretical discussion, we model the probability to observe a particular value of a bank's z-score Z and, therefore, its probability of default, as a function of its derivatives use. Assuming that there are two classes of banks that use derivatives – hedgers and traders – then the probability of observing $Z = z_{i,t}$ is given by

$$\begin{aligned} Prob(Z = z_{i,t}) = Prob(Z = z_{i,t} \; | Hedger_{i,t}) Prob(Hedger_{i,t}) + \\ Prob(Z = z_{i,t} \mid Trader_{i,t}) Prob(Trader_{i,t}) \end{aligned} \tag{3.13}$$

where $Prob(Hedger_{i,t}) + Prob(Trader_{i,t}) = 1$. If a bank's derivatives use is perfectly observed, we can set the probabilities $Prob(Hedger_{i,t})$, $Prob(Trader_{i,t})$ to one or zero respectively depending on the derivatives use that is observed. If, on the other hand, how derivatives are used is not perfectly observed, we have $Prob(Hedger_{i,t})$, $Prob(Trader_{i,t}) \geq 0$. Since default risk decreases as banks hedge more, traders should have a low z-score with a higher probability than hedgers and vice versa.

Modeling the distribution of z-scores in this way allows us to back out the probability of a bank being a hedger or a trader conditional on its current z-score simply by applying

Bayes' theorem to obtain

$$Prob(Hedger_{i,t} \mid Z = z_{i,t}) = \frac{Prob(Z = z_{i,t} \mid Hedger_{i,t})Prob(Hedger_{i,t})}{Prob(Z = z_{i,t})} \tag{3.14}$$

$$\begin{aligned} Prob(Trader_{i,t} \mid Z = z_{i,t}) &= \frac{Prob(Z = z_{i,t} \mid Trader_{i,t})Prob(Trader_{i,t})}{Prob(Z = z_{i,t})} \\ &= 1 - Prob(Hedger_{i,t} \mid Z = z_{i,t}) \end{aligned} \tag{3.15}$$

We classify banks based on the estimates of these probabilities using a naïve Bayes classification, i.e., a bank i in quarter t is classified as a hedger if $Prob(Hedger_{i,t} \mid Z = z_{i,t}) > Prob(Trader_{i,t} \mid Z = z_{i,t})$.

To estimate $Prob(Hedger_{i,t} \mid Z = z_{i,t})$ and $Prob(Trader_{i,t} \mid Z = z_{i,t})$, we specify a latent class regression model by writing the logarithm of the z-score as a linear function of the bank characteristics separately for hedgers and traders:

$$log(z_{i,t}) = \begin{cases} \beta'_{Hedger}X_{i,t} + \varepsilon_{i,t,Hedger} \\ \beta'_{Trader}X_{i,t} + \varepsilon_{i,t,Trader} \end{cases} \tag{3.16}$$

where $X_{i,t}$ is a vector containing a constant and a set of quarterly demeaned bank characteristics, and β_{Hedger} and β_{Trader} are vectors of class-specific regression coefficients. For the error term, we have $\varepsilon_{i,t,c} \sim \mathbb{N}(0, \sigma_c^2)$ with $c = \{Hedger, Trader\}$, i.e., the error terms are independently but not identically distributed across classes, since we assume normally distributed error terms with zero-mean values, but with class-specific variances. The normality of the error term is crucial to ensure the identifiability of the latent class regression model (see Grün and Leisch (2008)).

Using the notation in Equation (3.16), we can recast the assumption that, everything else equal, hedgers should be less risky than traders in terms of the expected log z-score for hedgers and traders:

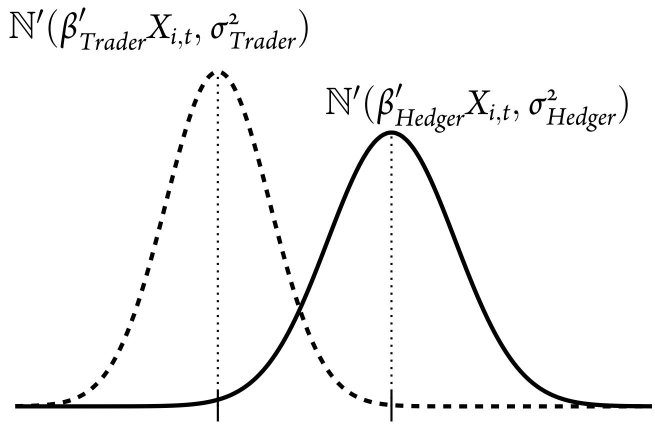
$$\mathbb{E}(log(z_{i,t}) \mid X_{i,t}, Hedger) = \beta'_{Hedger}X_{i,t} > \mathbb{E}(log(z_{i,t}) \mid X_{i,t}, Trader) = \beta'_{Trader}X_{i,t} \tag{3.17}$$

This means that, controlling for bank characteristics, the expected log z-score of the bank that is engaging in hedging should be larger than the expected log z-score of the bank that is engaging in trading. Furthermore, by using the notion of the regression model in Equation (3.16), we can now easily write the parametric forms for $Prob(Z = z_{i,t} \mid Hedger_{i,t})$ and $Prob(Z = z_{i,t} \mid Hedger_{i,t})$ as

$$Prob(Z = z_{i,t} \mid Hedger_{i,t}) = \tag{3.18}$$
$$= f_{Hedger}(log(z_{i,t}) \mid X_{i,t}, \beta_{Hedger}, \sigma_{Hedger}) = \mathbb{N}(\beta'_{Hedger}X_{i,t}, \sigma^2_{Hedger})$$
$$Prob(Z = z_{i,t} \mid Trader_{i,t}) = \tag{3.19}$$
$$= f_{Trader}(log(z_{i,t}) \mid X_{i,t}, \beta_{Trader}, \sigma_{Trader}) = \mathbb{N}(\beta'_{Trader}X_{i,t}, \sigma^2_{Trader})$$

Thus, the positions of the distributions $Prob(Z = z_{i,t} \mid Hedger_{i,t})$ and $Prob(Z = z_{i,t} \mid Trader_{i,t})$ depend on the expected z-score of a bank given its characteristics as shown in Figure 3.A.1. Using these distributions, we can, for each realized z-score, compute the probability of observing such a value under $\mathbb{N}(\beta'_{Hedger}X_{i,t}, \sigma^2_{Hedger})$ and under $\mathbb{N}(\beta'_{Trader}X_{i,t}, \sigma^2_{Trader})$. Equation (3.17) implies that a large z-score, i.e., the low probability of default, is most likely generated by $\mathbb{N}(\beta'_{Hedger}X_{i,t}, \sigma^2_{Hedger})$ than by $\mathbb{N}(\beta'_{Trader}X_{i,t}, \sigma^2_{Trader})$.

Figure 3.A.1: The Generic z-Score Distributions of Hedgers and Traders



Lastly, following the approach in Bos, Economidou, and Koetter (2010a) and Bos, Economidou, Koetter, and Kolari (2010b), we also parametrize the probabilities $Prob(Hedger_{i,t})$ and $Prob(Trader_{i,t})$ using the quarterly demeaned amounts of derivatives

that a bank uses and a constant, i.e.,
$D_{i,t} = (1, \textit{Interest Rate Derivatives}_{i,t}, \textit{Foreign Exchange Derivatives}_{i,t}, \textit{Credite Derivatives}_{i,t})$ as

$$Prob(Trader_{i,t}) = \omega_{Trader}\left(D_{i,t}, \eta\right) = \frac{exp\left(D'_{i,t}\eta\right)}{1 + exp\left(D'_{i,t}\eta\right)} \tag{3.20}$$

$$Prob(Hedger_{i,t}) = \omega_{Hedger}\left(D_{i,t}, \eta\right) = 1 - \omega_{Trader}\left(D_{i,t}, \eta\right) \tag{3.21}$$

Using Equations (3.16) to (3.21), we can write the distribution of the z-score $Prob(Z = z_{i,t})$ defined in Equation (3.13) in parametric form as

$$\begin{aligned} Prob(Z = z_{i,t}) =& g\left(log(z_{i,t}) \mid X_{i,t}, \beta_c, \sigma_c^2\right) \\ =& \sum_{c \in \{Hedger, Trader\}} \omega_c\left(D_{i,t}, \eta\right) f_c\left(log(z_{i,t}) \mid X_{i,t}, \beta_c, \sigma_c\right) \end{aligned} \tag{3.22}$$

Accordingly, we can rewrite Equations (3.14) and (3.15) in terms of these parametric distributions

$$\begin{aligned} & Prob(Hedger_{i,t} \mid Z = z_{i,t}) = \\ & = p_{i,t,Hedger} = \frac{f_{Hedger}(log(z_{i,t}) \mid X_{i,t}, \beta_{Hedger}, \sigma^2_{Hedger})\omega_{Hedger}\left(D_{i,t}, \eta\right)}{g\left(log(z_{i,t}) \mid X_{i,t}, \beta_c, \sigma_c^2\right)} \end{aligned} \tag{3.23}$$

$$Prob(Trader_{i,t} \mid Z = z_{i,t}) = p_{i,t,Trader} = 1 - p_{i,Hedger} \tag{3.24}$$

Estimation and Classification

Since all elements in Equation (3.22) have known parametric expressions, it is easy to write its log-likelihood function as

$$\hat{\theta} = arg\,max\left[\mathcal{L}_1 = \sum_{i \in I, t \in T} \sum_{c \in C} p_{i,t,c} log\left(\omega_c\left(D_{i,t}, \eta\right) f_c\left(log(z_{i,t}) \mid X_{i,t}, \theta_c\right)\right)\right] \tag{3.25}$$

$$\hat{\eta} = arg\,max\left[\mathcal{L}_2 = \sum_{i \in I, t \in T} \sum_{c \in C} p_{i,t,c} log\left(\omega_c\left(D_{i,t}, \eta\right)\right)\right] \tag{3.26}$$

where $C = \{Hedger, Trader\}$, I denotes the number of individual banks, T denotes the number of quarters, and $\theta = (\beta_{Hedger}, \beta_{Trader}, \sigma^2_{Hedger}, \sigma^2_{Trader})$. The coefficient vectors $\hat{\theta}$ and $\hat{\eta}$ can be found by using an expectation-maximization algorithm (see, e.g., Bos et al. (2010a), Bos et al. (2010b), Grün and Leisch (2008), Do and Batzoglou (2008)):

1. Set $p_{i,t,Hedger}$ and $p_{i,t,Trader} = 1 - p_{i,Hedger}$ for each observation to their initial values,
2. Estimate coefficient vectors $\hat{\theta}$ and $\hat{\eta}$ from Equations (3.25) and (3.26),
3. Update $p_{i,Hedger}$ and $p_{i,Trader}$ in Equations (3.23) and (3.24) using the estimated coefficient vectors $\hat{\theta}$ and $\hat{\eta}$ from step 2, and
4. Iterate between steps 2. to 3. until the sum of the log-likelihoods $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$ converges.

Let $\hat{\theta}^*$ and $\hat{\eta}^*$ be the estimated coefficient vectors at the maximum sum of the log-likelihoods $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$. We evaluate for each bank i in each quarter t the probability of it being a hedger or a trader $p^*_{i,t,Hedger}, p^*_{i,t,Trader}$ at $\hat{\theta}^*$ and $\hat{\eta}^*$. If $p^*_{i,t,Hedger} > p^*_{i,t,Trader}$, bank i is classified as a model-based hedger in quarter t and vice versa.

The outlined Expectation-Maximization (EM) algorithm tends to converge slowly and only to a local maximum (see Grün and Leisch (2008)). Several variants of the algorithm have been suggested in the literature to improve its performance. While the regular EM algorithm uses the probabilities $p_{i,t,Hedger}, p_{i,t,Trader}$ from the previous iteration in the maximizations in Equations (3.25) and (3.26), the Classification-Expectation-Maximization (CEM) algorithm assigns each observation to only one class. In other words, the probabilities $p_{i,t,Hedger}$ and $p_{i,t,Trader}$ from the previous iteration of the algorithm are not used as weights in Equations (3.25) and (3.26). Instead, $p_{i,t,Hedger}$ and $p_{i,t,Trader}$ are set exactly to one or zero for each observation depending on the maximum $p_{i,t,Hedger}$ and $p_{i,t,Trader}$ before carrying out the maximizations in Equations (3.25) and (3.26) for the next iteration. The CEM algorithm can be expected to exhibit better convergence behavior than the EM algorithm but does not yield maximum likelihood coefficient estimates as it maximizes the completed likelihood. However, we use the CEM algorithm to initialize the EM algorithm to mitigate the risk of convergence to a local maximum (see Grün and Leisch (2008), Biernacki, Celeux, and Govaert (2003), Karlis and Xekalaki (2003)). Specifically, we conduct 100 short runs (i.e., runs with ≤ 15 iterations) of the CEM algorithm. For each of these runs, we randomly initialize $p_{i,t,Hedger}$ and $p_{i,t,Trader}$. From these short runs of the CEM algorithm, we then pick the run with the highest sum of the log-likelihoods $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$ and use its probabilities $p_{i,t,Hedger}$ and $p_{i,t,Trader}$ to initialize the regular EM algorithm that is described above. Note that this approach does not use any information about how a bank uses its derivatives at any point in the algorithm. Only the total proportion of derivatives that a bank uses is taken as an input for the algorithm, but not how the bank reports on the use of these derivatives.

The model setup described so far can also be used to obtain estimates for the class-wise regression model in Equation (3.16) but using the observed supervision-based classification. In this case, we do not need to use the EM algorithm, since we observe whether a bank is hedger or a trader in any quarter, and we can simply estimate Equation (3.22) using the maximum likelihood estimation to obtain coefficient estimates for the regression model in Equation (3.16).

3.A.2 Variable Definitions

(a) *Total Assets* is the book value of assets.

(b) *Size* is the logarithm of the market value of equity divided by the logarithm of the cross-sectional mean value of the market value of equity. The market value of equity in each quarter is the within-quarter average of the product of daily stock prices and numbers of shares outstanding.

(c) *Leverage* is the difference between the book value of assets and the book value of (unweighted) equity, plus the market value of equity, divided by the market value of equity (see Acharya, Brownlees, Engle, Farazmand, and Richardson (2013), Mayordomo et al. (2014)).

(d) *Book-to-Market Ratio* is defined as the book value of equity divided by the market value of equity. The market value of equity for any specific quarter is computed as the time-series mean value over the weekly market values of equity. We take the logarithm of the Book-to-Market Ratio in all regressions since its distribution is skewed.

(e) *Intangible Assets* is computed as the sum of goodwill and trademarks, trade names, franchises, mortgage servicing rights, organization costs that would be amortized over a period of 60 months or more, and any other identifiable intangibles, divided by the Total Assets.

(f) *Foreign Deposits* is computed as the USD amount of all deposits held in offices outside the US and those held by foreign subsidiaries located outside of the US that are included in the consolidated accounting statement, divided by the Total Assets.

(g) *Foreign Currency Assets* are the sum of the fair values of all securities denominated in foreign currency, cash balances held outside of the US, loans granted to foreign governments and official institutions, and loans to foreign banks, divided by the Total Assets.

(h) *Non-Performing Loans* are the sum of all loans that are past due 90 days or more and still accruing and all non-accruing loans, divided by the Total Loans.

(i) *Total Loans* is the sum of all outstanding loans.

(j) *Loans-to-Deposits Ratio* is the ratio of total loans to total deposits.

(k) *Loan Loss Reserve* is the sum loans net of unearned income and lease financing receivables net of unearned income minus the sum of allowance for loan and lease losses and the allocated transfer risk reserve, divided by the Total Loans.

(l) *Demand Deposits* includes all regular checking accounts divided by Total Assets.

(m) *ROA* is computed as net income, including other comprehensive income, divided by the Total Assets.

(n) *Total Derivatives* is the sum of the gross notional amounts of Interest Rate-, Foreign Exchange-, and Credit-related derivatives, divided by the Total Assets.

(o) *IR-Derivatives* are the gross notional amounts of Interest Rate-related derivatives divided by the Total Assets.

(p) *FX-Derivatives* are the gross notional amounts of Foreign Exchange-related derivatives divided by the Total Assets.

(q) *Credit-Derivatives* are the gross notional amounts of Credit-related derivatives divided by the Total Assets.

3.A.3 Merton's Distance-to-Default

The Distance-to-Default measure suggested by Merton (1974) follows the same idea as the Z-score but is based on real options theory. Hence, its basis is that the equity owners of the bank have a put option with the value of debt being the strike price. The Merton (1974) Distance-to-Default is, then, the point at which it is ideal for the owners to exercise their option. Most model inputs for the Distance-to-Default measure cannot be observed; rather, they need to be estimated. To compute the measure, we follow the approach proposed by Bharath and Shumway (2008) and compute the Distance-to-Default as

$$DD_{it} = \frac{log\left(VA_{it}/D_{it}\right) + \left(r_t - 0.5\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$$

where VA_{it} and D_{it} are the market value of the total assets and the two-year moving average of the book value of total liabilities respectively. σ_A is the volatility of the total assets and r_t is the risk-free interest rate. T denotes the time to maturity of the real option. We estimate $VA_{it} = VE_{it} + D_{it}$. VE_{it} is the mean market value of equity in quarter t computed as the mean over the daily closing mid-prices of the bank's stocks and the shares outstanding. D_{it} is the trailing two-year moving average of the book value of debts. r_t is the 90-day T-bill rate, and T is one quarter. We estimate σ_A as

$$\sigma_A = \frac{VE}{VA}\sigma_E + \frac{D}{VA}\sigma_D$$

, with $\sigma_D = 0.05 + 0.25\sigma_E$, and σ_E is the standard deviation of the market value of equity within the current quarter.

3.A.4 Summary Statistics

Table 3.A.1: Summary Statistics – Classification Variables

Notes: In this table, we provide summary statistics for all variables used in the estimation of class membership probabilities. The sample consists of 12,593 quarterly observations of 454 banks from QIV:1997 to QIV:2015. Detailed variable definitions are provided in Appendix 3.A.2.

Statistic	Mean	St. Dev.	25% Pctl.	Median	75% Pctl.
log(Z-score)	4.762	1.193	4.133	4.942	5.632
Size	0.898	0.134	0.800	0.881	0.975
Leverage	0.867	0.061	0.834	0.869	0.904
log(Book-to-Market)	−0.275	0.610	−0.689	−0.334	0.036
Intangible Assets	0.019	0.018	0.004	0.014	0.028
ROA (in %)	0.192	0.376	0.154	0.243	0.319
Derivatives	0.673	3.333	0.012	0.039	0.130
IR-Derivatives	0.539	2.736	0.010	0.036	0.118
FX-Derivatives	0.108	0.544	0.000	0.000	0.0002
Credit-Derivatives	0.022	0.195	0.000	0.000	0.000
Non-Performing Loans	0.017	0.021	0.005	0.010	0.021
Loan Loss Reserve	0.016	0.008	0.011	0.014	0.018
Loans-to-Deposits	0.902	0.198	0.800	0.909	1.005
Foreign Currency Deposits	0.025	0.094	0.000	0.000	0.000
Foreign Currency Assets	0.006	0.023	0.000	0.000	0.0001
Demand Deposits	0.087	0.065	0.039	0.071	0.119

4

Nobody Knew That Measurement Error Could Be So Complicated: A Note on Estimating Betas and Market Risk Premiums[1]

4.1 Introduction

The estimation of betas is the basis of many applications in finance. From a theoretical perspective, researchers are interested in explaining the cross-sectional variation in stock returns. Building on that interest, practitioners use asset pricing theory to determine the costs of capital for budgeting decisions, to develop investment strategies, and to analyze the

[1]This chapter is based on a working paper co-authored by Paulo Rodrigues (Maastricht University) and Rogier Quaedvlieg (Erasmus University Rotterdam).

performance of mutual or hedge funds. The workhorse model provided to practitioners is the Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965). Welch (2009) finds that 75% of finance professors recommend using the CAPM and Graham and Harvey (2001) find that 73.5% of CFOs follow this advice. Indeed, Fama and French (2003) and Morey (2015) point out that the CAPM is an integral part of finance education, since it is often the only asset pricing model taught in MBA investment courses and all four of the best-selling introductions to finance textbooks in 2015 extensively cover the CAPM.

Notwithstanding the CAPM's popularity, the model performed poorly in empirical tests as one of its key predictions – a positive premium for holding market risk (beta) – has repeatedly been rejected by researchers. For example, Black, Jensen, and Scholes (1972), Fama and MacBeth (1973), and Fama and French (1992) find small positive market risk premiums that are statistically insignificant.[2] Haugen and Heins (1975), Campbell (1987), and Nelson (1991) even report a negative market risk premium. These results have led to the conventional wisdom that there is no positive or even a negative trade-off between market risk (beta) and average returns, as investors do not receive a premium for holding market risk. This anomaly (usually referred to as the low-risk anomaly or the beta anomaly) is commonly explained using a mix of behavioral biases and market frictions.[3]

In this chapter, we show that this conventional wisdom is flawed since flat or negative market risk premiums can be attributed to statistical bias rather than behavioral bias or market frictions. We argue that betas are difficult to estimate precisely, since estimates of betas are impacted by three sources of statistical biases: price measurement error, sampling error, and time-series variation in betas.[4] If betas are estimated using OLS, the measurement error in the underlying return data due to price staleness and illiquidity, the sampling bias due to the small size of the samples for estimation, and the time-series variation in true betas all lead to noisy beta estimates. The interrelation between these errors produces non-trivial trade-offs between bias and variance in the estimated betas. We

[2]Similar results are reported in French, Schwert, and Stambaugh (1987), Baillie and DeGennaro (1990), and Campbell and Hentschel (1992).

[3]Ang (2014) and Baker, Bradley, and Wurgler (2011) provide excellent overviews on the different explanations.

[4]Each of these statistical biases has been discussed in isolation in various contexts, but their combined effect on the estimation of market risk premiums has not been considered. See, e.g., Blume (1975), Scholes and Williams (1977), Dimson (1979), Blume and Stambaugh (1983), Jagannathan and Wang (1996), Asparouhova, Bessembinder, and Kalcheva (2010), Asparouhova, Bessembinder, and Kalcheva (2013).

demonstrate analytically that, as a consequence of these statistical biases, the standard two-pass Fama and MacBeth (1973) procedure to estimate the market risk premium yields heavily attenuated estimates and biased statistical inferences. The issue of noisy beta estimates and the resulting errors-in-variables problem in the application of the standard two-pass Fama and MacBeth (1973) procedure is well known (see, e.g., Kim (1995)). However, our results suggest that attenuation bias in market risk premiums in the presence of price measurement error, sampling error, and time-series variation in betas increases attenuation bias. We are also able to provide conditions under which estimated market risk premiums can even become negative when the true premium is positive.

Consistent with our analytical propositions, we find significant and positive market risk premiums after minimizing the sampling error through more frequent sampling, minimizing the error due to time-series variation by estimating betas over small windows, and minimizing price measurement errors by maintaining a sample of liquid stocks. We find highly significant market risk premiums between approximately 3.9% and 8.3% per annum over the long term from 1926 to 2013. For a more recent sub-period, from 1999 to 2013, we also find positive and significant market risk premiums in a similar range, albeit with lower significance. We find p-values between 0.056 and 0.059 for the sub-period from 1999 to 2013 and p-values < 0.05 for the long term from 1926 to 2013. Our results suggest that, especially in the long term, sample price measurement errors in underlying return data has a large impact on estimated market risk premiums. In more recent sample periods, time-series variation seems to have become more important, likely reflecting changing market structures.

Our results are consistent with a growing body of literature questioning the empirical finding of a flat relationship between market risk (beta) and average returns. For example, Ghysels, Santa-Clara, and Valkanov (2005) find a significant positive market risk premium using a Mixed-Data Sampling (MIDAS) approach to estimate betas. Buss and Vilkov (2012) use option-implied variances and covariances to estimate betas and also find a significant positive market risk premium. Cosemans, Frehen, Schotman, and Bauer (2016) find a significant positive market risk premium using a Bayesian shrinkage estimator of betas. Furthermore, Bali, Engle, and Tang (2017) estimate a statistically significant positive market risk premium using time-varying conditional betas estimated using the DCC method of Engle (2002). All of these studies have in common that they employ sophisticated econometric techniques rather than simple OLS to increase the precision of estimated betas. However, in this chapter, we rely entirely on OLS estimation to ensure that our estimated betas retain the property of simply being the covariance between a market portfolio and an asset divided by the market portfolios variance, without imposing further assumptions.

For example, the option-implied betas of Buss and Vilkov (2012) are estimated using the risk-neutral measure rather than the physical probability measure. This choice could affect the estimation of market risk premiums derived from their betas, since betas estimated using the risk-neutral measure can be different from those estimated using the physical measure. If betas are estimated using the Bayesian shrinkage approach in Cosemans et al. (2016), they are not necessarily the ratio of covariance between market and asset return and the variance of the market. Furthermore, beta estimation approaches such as the MIDAS method in Ghysels et al. (2005), DCC methods in Bali et al. (2017) or the approach in Cosemans et al. (2016) impose strong structural assumptions.

The importance of the estimation technique in the context of the estimation of betas and market risk premiums has already been documented in numerous studies. For example, Glosten, Jagannathan, and Runkle (1993), Harvey (2001), and Turner, Startz, and Nelson (1989) each find both negative and positive market risk premiums depending on the employed estimation method. Most recently, Shanken and Zhou (2007) demonstrates that the Fama-MacBeth procedure for testing CAPM implemented by OLS yields accurate point estimates but also standard errors that are too large. Meanwhile, for example, GLS produces smaller standard errors but also biased point estimates.

Our findings suggest that the sensitivity of OLS to the properties of the underlying return data significantly influences the outcome of the standard Fama-MacBeth procedure, leading to the appearance of spurious asset pricing "anomalies". Especially for recent time periods, OLS should only be applied for sufficiently large samples of liquid stocks, and if high sampling frequencies are available, short estimation windows should be used. Consistent with this argument, Grauer and Janmaat (2009) and Connolly and Rendleman (2009) demonstrate in extensive simulations that, under idealized conditions, the Fama-MacBeth procedure for testing CAPM yields reasonably accurate point estimates of the market risk premium. However, in their simulations, standard deviations often remain too large to reliably reject the hypothesis that the market risk premium is zero even if the true market risk premium is positive. As conditions become less ideal, point estimates tend to become increasingly biased. Hence, even if the true market risk premium is positive and the data is generated by a CAPM-like model, standard implementations of the Fama-MacBeth procedure may lack the statistical power to reliably reject the hypothesis that the market risk premium is zero due to the noisy beta estimates.

Estimating betas and market risk premiums precisely has consequences beyond academic discussions. The idea of a negative or flat market risk premium due to behavioral biases and market frictions has inspired the development of trading strategies to exploit the alleged inefficiency in the market. In particular, Frazzini and Pedersen (2014) coined the term "betting-against-beta" to describe trading strategies that take long positions in low-beta stocks and short positions in high-beta stocks. If the market risk premium is positive, an investor can only realize above-average returns by taking above-average market risk. However, if there is a negative market risk premium, the opposite would be true. Investors who invest in high-beta stocks would realize below-average returns, investors who invest in low-beta stocks realize above-average returns, and betting-against-beta trading strategies would be profitable. We extent our empirical analysis to investigate the relationship between our market risk premium estimates and the corresponding average returns of a simple betting-against-beta trading strategy. Unsurprisingly, whenever we find positive market risk premiums, average betting-against-beta returns are negative and vice versa. As we exclude illiquid stocks from our sample, estimated market risk premiums increase in magnitude and significance. At the same time, average betting-against-beta returns decrease from positive to negative. Our findings regarding betting-against-beta trading strategies are consistent with the results documented in Novy-Marx and Velikov (2016) and Li, Sullivan, and Garcia-Feijoo (2014), which show that the performance of betting-against-beta strategies are concentrated in small and illiquid stocks. This suggests that the good past performance of betting-against-beta strategies was driven by an illiquidity premium rather than by the exploitation of an inefficiency in the market.

Of course, our approach concerning statistical biases, restricting our sample to those stocks that are the least affected by the statistical biases, comes at a cost: Restricting the sample used for the estimation of market risk premiums in a non-random way limits the interpretation of our findings. We show that the OLS beta estimation and the associated market risk premium perform well if they only consider stocks for which the discussed statistical biases are not present. The estimation approach suggested in Cosemans et al. (2016) takes a different perspective on the issue of statistical biases in the data by adjusting the estimator so that it can use an unrestricted sample. However, adjusting the estimator comes at the cost of the estimated betas no longer being a ratio of covariance and variance. Our approach of restricting the sample betas maintains this basic property, as they are still estimated by means of OLS, but our resulting market risk premiums are no longer universal.

However, the lack of universality of our estimated market risk premiums is not necessarily problematic. In a recent paper, Patton and Weller (2018) show that market risk premiums are not constant in the cross-section, and different subsets of stocks have different market risk premiums. We argue that while it is possible to identify the market risk premium properly using simple OLS for some of these subsets, doing so is not possible for other subsets. Our results suggest that whether or not the market risk premium of a particular subset of the market can be properly identified by OLS estimation depends largely on the degree of statistical biases in the data. In those subsets in which we cannot properly identify market risk premiums by OLS estimation, other econometric techniques should be employed for estimation.

4.2 Beta, Risk Premiums and Measurement Error

The most popular approach for testing whether a risk factor receives a premium in the cross-section of stock returns is the two-pass regression procedure suggested by Fama and MacBeth (1973). In our context, we aim to estimate the premium for market risk as measured by a simple single market factor beta. The Fama and MacBeth (1973) procedure comprises the OLS estimation of betas from a time series of asset returns and market index returns in a first-stage regression, yielding a cross-section of individual asset betas. In the second stage, the cross-section of realized returns is regressed on the cross-section of these estimated asset betas. The market risk premium is the estimated OLS coefficient from this second-stage regression.[5] In this sub-chapter, we derive the properties of the first-stage

[5] There are several standard procedures to address the generated regressors issue in this context. Specifically, asset betas can be estimated on rolling-windows to create a panel of asset betas rather than a single cross-section. The second stage is than applied to the single cross-sections defined by the time periods of the sample. This results in a time series of market risk premiums. The final estimate of the market risk premium is then obtained by computing the average over the time series of the market risk premium. In our empirical

OLS beta estimator in the presence of price measurement error, sampling error, and time-series variation in true betas. Building thereon, we derive the properties of the second-stage OLS market risk premium estimator.

To derive a more formal definition of the problem at hand, we assume that returns are generated by a CAPM-type model in the form

$$r_{t,i} = \beta_{t,i} f_t + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \sim (0, \sigma^2) \tag{4.1}$$

where $r_{t,i}$ and f_t denote the excess log returns of an individual asset i and the market factor respectively. The error term $\varepsilon_{t,i}$ is assumed to be homoscedastic for the sake of simplicity, but this assumption could easily be relaxed. To further simplify the notation, we omit the i subscript where it is not necessary for the sake of clarity.[6] In the first stage, we aim to estimate the expected market beta factor $\beta_T = \mathbb{E}\left(\beta_T \mid r_{T_0}, \ldots, r_T, f_{T_0}, \ldots, f_T\right)$ using a rolling or expending estimation window $[T_0, T]$ of asset and market factor returns. Given a cross-section of betas, Fama and MacBeth (1973) show that the market risk premium can be estimated using the following cross-sectional regression:

$$R_{T,i} = \lambda_T \beta_{T,i} + \tilde{\varepsilon}_{T,i}, \quad \tilde{\varepsilon}_{T,i} \sim (0, \sigma^2_{\tilde{\varepsilon}}) \tag{4.2}$$

where $R_{T,i}$ represents the cross-section of single stock simple returns in excess of the risk-free rate. Since $\beta_{T,i}$ is unobservable in applications, it is replaced by its estimate $\hat{\beta}_{T,i}$, obtained from Equation (4.1).

Estimation of betas in this setting is complicated by the well-known result that betas are not constant but vary over time (see Jagannathan and Wang (1996), Fabozzi and Francis (1978), Rosenberg and Guy (1976), Blume (1975)). Put simply, time-series variation in betas can be captured by the notion of a structural break, i.e.,

$$\beta_t = \begin{cases} \beta_0 \text{ if } t < T_1 \\ \beta_1 \text{ if } t \geq T_1 \end{cases} \tag{4.3}$$

analysis, we follow this approach.

[6]Since Equation (4.1) defines time-series regressions that are performed "asset-by-asset," the i subscript can be omitted there without a loss of generality.

Hence, if β_T is estimated over any time period $[T_0, T]$ with $T_0 < T_1 < T$, the resulting estimate $\hat{\beta}_T$ is biased due to the structural break. Moreover, the estimated beta will converge to a weighted average of β_0 and β_1. In the second stage of the Fama-MacBeth procedure, returns $R_{T,i}$ are regressed on estimates of $\hat{\beta}_{T,i}$, but due to the break, we have $\beta_{T,i} = \beta_{1,i}$. Therefore, the more observations prior to the break that are included in the estimation window, the greater the bias in $\hat{\beta}_{T,i}$.

Such a bias can be avoided by estimating betas over shorter time periods, i.e., by picking a starting point for the estimation window $T_0 \geq T_1$. However, for a given frequency of returns, the sample size declines the shorter the estimation window; thus the variance of the estimator increases. In some situations, it may be preferable to choose $T_0 < T_1$, i.e., deliberately including the breakpoint in the estimation window, if the larger sample size sufficiently reduces the variance of the estimator (see Pesaran and Timmermann (2007)). Thus, time-series variation in betas introduces a trade-off between the bias and variance of the beta estimator.

Rather than keeping the frequency of returns fixed, it is possible to increase the frequency of returns within a shorter estimation window and therefore increase the sample size. For example, if we reduce the estimation window from 12 months to six months because we suspect a structural break, we could use daily rather than weekly returns to estimate betas. However, increasing the frequency of returns has its own set of problems.

Observed asset returns may suffer from stale prices and illiquidity, leading to a situation in which we cannot observe the true returns (see Asparouhova, Bessembinder, and Kalcheva (2013), Asparouhova, Bessembinder, and Kalcheva (2010), Blume and Stambaugh (1983), Dimson (1979), Scholes and Williams (1977)). We define price staleness as a situation in which observed prices remain sticky for a fraction of the interval between returns. This fraction determines how much returns deviate from their actual, latent returns over that time period. Thus, while the true return is $r_t = p_t - p_{t-1}$, where p_t is the log-price, we observe $\tilde{r}_t = p_{t-k} - p_{t-1-k}$, with $0 \leq k < 1$. If $k = 0$, then the observed return equals the true return, i.e., $\tilde{r}_t = r_t$. However, if $k > 0$, the estimates of the market beta are biased.[7] If returns are uncorrelated over time, we have $Corr(r_t, \tilde{r}_t) = 1 - k$.

If there is price staleness, then the degree to which it affects the first and second stage

[7] Of course, the fixed, constant shift of the return due to stale prices is a strong assumption that is purely made for the purposes of illustrating the effect of price staleness.

estimations of the Fama-MacBeth procedure depends on the frequency at which returns are observed, i.e the frequency of time steps in the estimation window. If k is large, then the biases due to stale prices will be greater if one-period returns are used compared to h-period returns, with $h > 1$ since the correlation between observed and true return converges to one as h increases. Let $r_t^{(h)}$ be the h-period return $p_t - p_{t-h} = \sum_{j=0}^{h-1} r_{t-j}$, then $Corr(r_t^{(h)}, \tilde{r}_t^{(h)}) = 1 - k/h$.

In summary, the issues of time-series variation, price measurement error, and sampling error are interconnected. In the presence of time varying betas short estimation windows of frequently sampled returns reduce the influence of time-series variation and sampling error. However, in the simultaneous presence of stale prices estimated betas will be biased. Mitigating this new bias by aggregating returns again reduces the sample size and thus would increase the estimator variance.

4.2.1 Implications For The First Stage

The first stage of the Fama-MacBeth procedure consists of time-series regressions of asset excess returns on market factor excess returns. $\beta_T = Cov_T(f_t, r_t)/Var_T(f_t)$, thus beta is the ratio of the covariance between the return and the market factor and the variance of the market factor at time T. In practice, we will always have to approximate this with a sample counterpart. Assume, for notational simplicity, that $\mathbb{E}(r_t) = \mathbb{E}(f_t) = 0$; moreover, let $\mathbb{E}(f_t^2) = \sigma_f^2$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2$.

To accommodate the choices regarding estimation window length and sampling frequency of returns we write the standard OLS estimator of beta as

$$\hat{\beta}_T(T_0, h) = \frac{\sum_{i=0}^{(T-T_0)/h} \left[\sum_{j=0}^{h-1} \tilde{r}_{T-ih-j} \sum_{j=0}^{h-1} f_{T-ih-j}\right]}{\sum_{i=0}^{(T-T_0)/h} \left[\sum_{j=0}^{h-1} f_{T-ih-j}\right]^2} \tag{4.4}$$

where h the degree of return aggregation, i.e. the larger h the lower the sampling frequency of returns. In terms of the often-used rolling-window estimation of betas T_0 constitutes the starting point of each window and $T - T_0$ the window length. For example, if we observe daily returns, $h = 20$ would amount to aggregating the daily returns to monthly returns, assuming 20 trading days in a month. Thus, $T_0 = T - 20 \times 60$ would amount to using a

60-month rolling window. Note that if $h = 1$, the above formula reduces to

$$\hat{\beta}_T(T_0, 1) = \frac{\sum_{i=0}^{(T-T_0)} \tilde{r}_{T-i} f_{T-i}}{\sum_{i=0}^{(T-T_0)} f_{T-i}^2}$$

Using the above introduced notation and definitions, we proceed by discussing bias and variance of $\hat{\beta}_T(T_0, h)$ for three different scenarios: betas are constant over time and there is no price staleness, time-varying betas but no price staleness, and price staleness but are constant over time.

The Ideal Case: $\beta_T = \beta_1 = \beta_0$ and $k = 0$

We assume that market betas are constant over time and prices are not stale such that the observed returns equal the true returns, i.e. $\beta_T = \beta_1 = \beta_0$ and $k = 0$ in the previously introduced notation. Hence, we have $\tilde{r}_t = r_t$ and $r_t = \beta_T f_t + \varepsilon_t$. Since the complications introduced in the previous sub-chapter are not present in this scenario the properties of the β_T estimator trivially follow from standard OLS theory, and therefore $\hat{\beta}_T(T_0, h) \rightarrow \beta_T$, for all T_0 and h. It is easy to see that the variance of the estimator is

$$Var(\hat{\beta}_T) = \frac{h\sigma^2}{(T - T_0)\sigma_f^2} \tag{4.5}$$

The denominator is independent of aggregation since $\mathbb{E}[(\sum_{j=0}^{h-1} f_{t+j})^2] = h\mathbb{E}(f_t^2)$. The numerator is independent of T_0, if we maintain the assumption of homoskedasticity of the error term and increasing in h. Based on this, it follows that the estimation strategy producing the most precise beta estimates would be to choose the longest estimation window and the highest possible sampling frequency of returns, i.e. no aggregation of returns to maintain the original data frequency, $h = 1$ and $T_0 \rightarrow -\infty$. Next we consider how deviations from this ideal case affect the properties of the first stage of the Fama-MacBeth procedure.

Time-varying Beta: $\beta_T = \beta_1 \neq \beta_0$ and $k = 0$

In this scenario we assume that prices are not stale and thus we observe the true returns $\tilde{r}_t = r_t$, but beta exhibits a structural break at time T_1 such that $\beta_T = \beta_1 \neq \beta_0$. If we choose the length of the estimation window $T - T_0$ such that $T_0 > T_1$ the estimator of β_T will be unbiased and the variance of the estimator is given by Equation (4.5).

However, in any practical application the time of the structural break is unknown. Hence, choosing T_0 introduces a trade-off between the potential bias due to the structural break if the estimation window is long and an increased variance of the estimator if the estimation window is short. If T_0 is too far in the past, $\hat{\beta}_T$ is likely to be severely biased, but its variance will be limited. A recent T_0 will lead to an estimate with small or no bias, but high variance. Both issues will affect the estimates of the market risk premium in the second stage of the Fama-MacBeth procedure.

If we choose $T_0 < T_1$ the resulting estimate of β_T will converge to a weighted average of the true betas before and after the structural break. To see this, we decompose the standard OLS estimator of β_T defined in Equation (4.4) into one part that uses data before the structural break and another part that uses data after the structural break:

$$\begin{aligned}\hat{\beta}_T(T_0, h) = & \frac{\sum_{i=0}^{(T-T_1)/h} \left[\sum_{j=0}^{h-1} r_{T-ih-j} \sum_{j=0}^{h-1} f_{T-ih-j}\right]}{\sum_{i=0}^{(T-T_0)/h} \left[\sum_{j=0}^{h-1} f_{T-ih-j}\right]^2} \\ & + \frac{\sum_{i=0}^{(T_1-T_0)/h} \left[\sum_{j=0}^{h-1} r_{T_1-ih-j} \sum_{j=0}^{h-1} f_{T_1-ih-j}\right]}{\sum_{i=0}^{(T-T_0)/h} \left[\sum_{j=0}^{h-1} f_{T-ih-j}\right]^2}\end{aligned} \tag{4.6}$$

The first term's numerator uses observation after the breakpoint T_1 and converges to β_1, whereas the second term's numerator only uses pre-break data and converges to β_0. As a result we have

$$\hat{\beta}_T(T_0, h) \to \frac{T - T_1}{T - T_0}\beta_T + \frac{T_1 - T_0}{T - T_0}\beta_0 = \nu_1\beta_T + \nu_0\beta_0 \tag{4.7}$$

Hence, while the choice of sampling frequency, i.e. the level of return aggregation h, is irrelevant in terms of bias, the length of the estimation window is crucial. The greater the fraction of total observations in the period prior to the breakpoint, the more the resulting estimate of beta will be biased towards the pre-breakpoint beta.[8] Next we turn our attention to the other part of the trade-off and show how the length of the estimation window affects the variance of the estimator under the current scenario.

The variance-bias trade-off is complicated since the magnitude of the structural break $\beta_1 - \beta_0$ has an impact on the variance of the estimator of β_T. The greater this magnitude is, the greater the resulting residual variance of the mis-specified model. To see this, we write

$$\begin{aligned} r_t &= \hat{\beta}_T f_t + u_t \\ u_t &= \varepsilon_t + (\beta_0 - \hat{\beta}_T) f_t I_{t<T_1} + (\beta_1 - \hat{\beta}_T) f_t I_{t\geq T_1} \end{aligned}$$

Here $I_{t<T_1}$ is an indicator variable which equals one for $t < T_1$ and zero otherwise and $I_{t\geq T_1}$ denotes accordingly an indicator variable which equals one for $t \geq T_1$ and zero otherwise. In this situation u_t will be heteroskedastic even if we maintain our assumption that $Var(\varepsilon_t) = \sigma^2$ and $Var(f_t) = \sigma_f^2$ are both constant. The variance of the estimator is then given by

$$Var(\hat{\beta}_T) = \frac{\sum_{i=0}^{(T-T_0)/h} \left[\sum_{j=0}^{h-1} f_{T-ih-j} u_{T-ih-j} \right]^2}{\left[(T - T_0)\sigma_f^2 \right]^2} \tag{4.8}$$

We can find a lower bound for $Var(\hat{\beta}_T)$ by imposing specific values for the weights v_0 and v_1 in equation (4.7). In particular, if we assume that $v_0 = v_1 = 1/2$ and therefore $\hat{\beta}_T \to (\beta_0 + \beta_1)/2$, then we have

$$Var(\hat{\beta}_T) \geq \frac{h\sigma^2 + h\sigma_f^2[(\beta_0^2 + \beta_T^2)/4 - (\beta_0\beta_T)/2]}{(T - T_0)\sigma_f^2} \tag{4.9}$$

Thus, the effect of the sampling frequency of returns, i.e. the level of return aggregation, on the level of $Var(\hat{\beta}_T)$ is not immediate. While the lower bound of the variance is increasing the level of aggregation h, the result still depends on the specific choice of T_0. This presents

[8] Of course, beta could, strictly speaking, be 'estimated' on a single observation, where we are sure to not be impacted by changes in the true parameter. However, the variance of such an estimator is prohibitive, and it is clear that it should be favorable to increase the estimation window to reduce the variance of the estimator.

a delicate bias-variance trade-off. $Var(\hat{\beta}_T)$ is increasing in T_0, thus the shorter the estimation window, the higher the variance of the beta estimator, but the lower the potential bias. To what degree the higher variance is associated with a shorter estimation window can be offset through less return aggregation, and thus higher return frequency still depends on the amount of pre-breakpoint observations included in the estimation window.

Price Staleness: $\beta_T = \beta_0 = \beta_1$ and $k > 0$

In this sub-chapter we assume that betas are constant over time, but asset returns suffer from price staleness, i.e. $\beta_T = \beta_0 = \beta_1$ and $k > 0$. This implies that there is no structural break in betas, but the observed returns deviate from the true returns. Thus $r_t \neq \tilde{r}_t = p_{t-k} - p_{t-1-k}$. Similar to Scholes and Williams (1977), we assume that the market factor is perfectly liquid and observable, but the last price we observe of our asset is due to a trade that occurred in the recent past, thus leading to a degree of price staleness and non-synchronicity between the market-factor returns and asset returns. The impact on the estimation of β_T depends on the severity of the price staleness, i.e. on the size of k. Recall that $Corr(r_t^{(h)}, \tilde{r}_t^{(h)}) = 1 - k/h$. Hence, by standard arguments the estimated beta converges to

$$\hat{\beta}_T(T_0, h) \to (1 - k/h)\beta_T \tag{4.10}$$

Since the observed returns deviate more from the 'true' return, the longer the stale prices last, i.e. the larger k, the greater the bias in the estimated beta. However, the problem can be alleviated by sampling at a lower frequency, i.e. by increasing h.

This, again, introduces a trade-off between bias and estimator variance. While increasing the level of return aggregation h reduces the bias due to price staleness. Doing so reduces the number of observations used in the estimation and thus increases the variance of the estimator.

The presence of price staleness raises two issues. First, we do not observe the true return. Second, asset returns, and market portfolio returns are not synchronous any more. The impact from the latter can be avoided by deviating from standard OLS estimation using approaches like the estimators suggested in Scholes and Williams (1977) or Dimson (1979). However, our goal is to remain in the framework of standard OLS estimation in both stages of the Fama-MacBeth procedure to derive its properties under common non-ideal conditions in the underlying data.

To derive the variance of the estimator we write

$$\tilde{f}_t = (1 - k/h)f_t + \sqrt{1 - (1 - k/h)^2} z_t \tag{4.11}$$

where $z_t \sim (0, \sigma_f^2)$ and $\tilde{f}_t$ is the corrected market portfolio return. Note that $\tilde{f}_t$ is not the 'true' stale market return, i.e. the market portfolio return time-matched to the observed returns, but rather is adjusted to have the appropriate variance and correlation with f_t. Hence, we have

$$\begin{aligned}
\tilde{r}_t &= \beta_T \tilde{f}_t + \tilde{\varepsilon}_t \\
&= \beta_T f_t + \beta_T(\tilde{f}_t - f_t) + \tilde{\varepsilon}_t \\
&= \hat{\beta}_T f_t + \underbrace{(\beta_T - \hat{\beta}_T)f_t + \beta_T(\tilde{f}_t - f_t) + \tilde{\varepsilon}_t}_{=u_t} \\
&= \hat{\beta}_T f_t + u_t
\end{aligned}$$

Using Equation (4.11) we can write the error term u_t as

$$\begin{aligned}
u_t &= (\beta_T - \hat{\beta}_T)f_t + \beta_T(\tilde{f}_t - f_t) + \tilde{\varepsilon}_t \\
&= \beta_T \sqrt{1 - (1 - k/h)^2} z_t + \tilde{\varepsilon}_t
\end{aligned}$$

Noting that $Var(\varepsilon_t) = Var(\tilde{\varepsilon}_t)$ we can now write the variance of $\hat{\beta}_T$ as

$$Var(\hat{\beta}_T) = \frac{h\sigma^2 - \sigma_f^2 \beta^2 [k^2/h - 2k]}{(T - T_0)\sigma_f^2} \tag{4.12}$$

Since under the current scenario we assumed beta to be constant over time, trivially the best choice would be to set the estimation window length $(T - T_0)$ as long as possible. The trade-off between bias and variance lies in the level of aggregation. As previously shown the bias due to stale prices is decreasing in h and the variance of the estimator is increasing in h. Thus, if returns suffer from stale prices, bias in the estimator of the market factor beta can be reduced by aggregating returns. Intuitively, a daily asset return will be affected to a greater extent if the observed price of the asset does not change over the course of two days due to a lack in trading activity than the monthly return. However, aggregation comes at the cost of a lower number of observations within the estimation window, and thus, a higher variance of the beta estimator.

4.2.2 Implications For The Second Stage

So far, we have formally defined bias-variance trade-offs in the first stage of the Fama-MacBeth procedure resulting from time-varying betas and price-staleness. We now discuss how these trade-offs affect the estimation of the market risk premium in the second stage.

The market risk premium defined on population quantities is given by

$$\lambda_T = Cov(R_{i,T}, \beta_{i,T})/Var(\beta_{i,T}) \tag{4.13}$$

Thus, the second stage of the Fama-MacBeth procedure is implemented by replacing the covariance and variance terms in (4.13) with their sample counterparts obtained from the time T cross-section of simple asset returns $R_{T,i}$ and estimated betas $\hat{\beta}_{T,i}$. Here we add an $i = 1, \ldots, N$ to the subscript to denote the individual assets, where N represents the number of individual assets in the dataset. We can write the estimated time T market risk premium using the regression:

$$R_{i,T} = \hat{\lambda}_T \hat{\beta}_{i,T} + \varepsilon_{i,T} \tag{4.14}$$

where $\hat{\lambda}_T$ is simply the standard OLS estimator

$$\hat{\lambda}_T = \frac{\sum_{i=1}^{N} R_{i,T}(\hat{\beta}_{i,T} - \bar{\beta}_T)}{\sum_{i=1}^{N}(\hat{\beta}_{i,T} - \bar{\beta}_T)^2} \tag{4.15}$$

with $\bar{\beta}_T = \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{i,T}$.

It is well understood that $\hat{\lambda}_T \to \lambda_T$ if β_T is measured without error. Unbiasedness or even consistency of the beta estimator alone are not sufficient to ensure consistency of the second stage. If estimated betas exhibit zero-mean measurement error, estimates of the market risk premium will be biased towards zero. If the estimated betas exhibit non-zero-mean measurement error the estimated market risk premium may converge to any arbitrary value. Thus, the properties of the first stage estimator are of crucial importance for the results of the second stage estimation.

We will now discuss the effect of structural breaks in betas and the effect of price-staleness on $\hat{\lambda}_T$. For notational convenience, we are going to summarize bias and variance by writing $\hat{\beta}_{i,T} = \beta_{i,T} + z_{i,T}$, where $z_{i,T}$ has mean $\mathbb{E}[\hat{\beta}_{i,T} - \beta_{i,T}]$ and variance $Var[\hat{\beta}_{i,T} - \beta_{i,T}]$. The exact expressions for the moments of $z_{i,T}$ per scenario have been derived in the previous sub-chapter. The classical measurement error assumption is an unbiased estimator $\mathbb{E}(z_{i,T}) = 0$, and that the measurement error is uncorrelated to all other variables $\mathbb{E}(R_{i,T}z_{i,T}) = \mathbb{E}(\beta_{i,T}z_{i,T}) = \mathbb{E}(\varepsilon_{i,T}z_{i,T}) = 0$. Under these assumptions $\hat{\lambda}_T$ is simply attenuated, i.e. biased towards zero.

As shown in the previous sub-chapter, the unbiasedness assumption implies that the estimation error averages to zero in the cross-section of stocks. This assumption is likely to be violated, especially in the presence of price staleness. The assumption that $\mathbb{E}(R_{i,T}z_{i,T}) = 0$ is also violated by construction since $R_{i,T}$ is one of the observations used for the estimation of beta, and its idiosyncratic component therefore partly drives $z_{i,T}$. However, this correlation will converge to zero when the estimation window increases in length. Moreover, this correlation is also alleviated when using $R_{i,T+1}$ on the left-hand side rather than $R_{i,T}$. Finally, the assumption that $\mathbb{E}(\beta_{i,T}z_{i,T}) = 0$ is also very likely violated as high 'current' betas were likely preceded by lower betas and vice versa. We will now discuss the same scenarios as in the previous sub-chapter in the same order.

Ideal Case: $\beta_T = \beta_1 = \beta_0$ and $k = 0$

In this scenario, betas are constant over time and returns do not suffer from price staleness. Hence, $\hat{\lambda}_T$ is only impacted by an attenuation bias due to sampling error, with the well-known result:

$$\hat{\lambda}_T \rightarrow \frac{Var(\beta_{i,T})}{Var(\beta_{i,T}) + Var(z_{i,T})}\lambda_T \tag{4.16}$$

Since we are not only interested in the potential bias of point estimates of the market risk premium but also in its statistical significance it is worthwhile to also consider the influence on statistical inference. For the estimator of the variance of $\hat{\lambda}_T$ we have the well known result

$$\hat{Var}(\hat{\lambda}_T) \to \Phi Var(\hat{\lambda}_T) + \Phi(1-\Phi)\lambda_T^2 \tag{4.17}$$

where $\Phi = Var(\beta_{i,T})/(Var(\beta_{i,T}) + Var(z_{i,T}))$. Therefore, not only the point estimate $\hat{\lambda}_T$ is biased but also the estimate of its variance $\hat{Var}(\hat{\lambda}_T)$. Unlike for the point estimate, the direction of the bias of $\hat{Var}(\hat{\lambda}_T)$ is unclear. The first term in the above equation suggests that the variance is downward-biased, but since the second term is positive the overall effect is unclear. However, the corresponding t-statistic will be biased downwards since

$$\frac{\hat{t}_{\hat{\lambda}}}{\sqrt{N}} = \frac{\hat{\lambda}_T}{\sqrt{\hat{Var}(\hat{\lambda}_T)}} \to \sqrt{\Phi}\frac{\lambda_T}{\sqrt{Var(\hat{\lambda}_T) + (1-\Phi)\lambda_T^2}} < \frac{\lambda_T}{\sqrt{Var(\hat{\lambda}_T)}} \tag{4.18}$$

Thus, if the attenuation bias is sufficiently large, point estimates of the market risk premium and the corresponding t-statistics will both be close to zero. Therefore, attenuation bias caused by noisy estimates of betas can, in principle, explain the finding of a flat and insignificant market risk premium reported in earlier studies. However, we have not yet established that it can lead to a negative estimate of market risk premium given that the true market risk premium is positive.

Since $Var(z_{i,T})$ is non-negative by definition, it is clear that an attenuating bias towards zero is unavoidable. However, feasible estimates of $Var(z_{i,T})$ can easily be obtained by means of the standard errors of the first-stage regressions, and therefore it is actually possible to de-bias the estimator, although this is rarely done in practice. Regardless, in this setup the bias in $\hat{\lambda}_T$ is clearly minimized, and asymptotically eliminated, by increasing the estimation window length. Hence, under the scenario of constant betas and no price staleness, the best strategy in the first stage estimation is to choose the longest possible estimation window and the highest possible return frequency.

Time-varying Beta: $\beta_T = \beta_1 \neq \beta_0$ and $k = 0$

As in the previous sub-chapter we will next consider the case of time-varying betas but continue to assume that returns are unaffected by stale prices. First, we consider the case of the measurement error being uncorrelated with beta, i.e., $\mathbb{E}(\beta_{i,T} z_{i,T}) = 0$. In this case we get the same expression for the attenuation bias as before:

$$\hat{\lambda}_T \rightarrow \frac{Var(\beta_{i,T})}{Var(\beta_{i,T}) + Var(z_{i,T})}\lambda_T \tag{4.19}$$

However, recall that in case of time-varying betas the variance of the estimator depends on the amount of observations before and after the breakpoint, and thus was given by:

$$\begin{aligned} Var(\hat{\beta}_{i,T}) &= \frac{\sum_{i=0}^{(T-T_0)/h}\left[\sum_{j=0}^{h-1} f_{T-ih-j}u_{T-ih-j}\right]^2}{\left[(T-T_0)\sigma_f^2\right]^2} \\ &\geq \frac{h\sigma^2 + h\sigma_f^2[(\beta_{i,0}^2 + \beta_{i,T}^2)/4 - (\beta_{i,0}\beta_{i,T})/2]}{(T-T_0)\sigma_f^2} \end{aligned}$$

where the bottom term is the variance assuming homoskedasticity, and therefore provides a lower bound. The lower bound is achieved when $\nu_0 = \nu_1 = 1/2$, i.e. if the breakpoint in betas is located in the middle of the estimation window and therefore $\beta_{i,T} = (\beta_{i,0} + \beta_{i,1})/2$. Furthermore, recall that $Var(\hat{\beta}_{i,T})$ under the constant beta assumption is simply $Var(\hat{\beta}_{i,T}) = h\sigma^2/(T-T_0)\sigma_f^2$ Hence, in case of time-varying betas the lower bound of $Var(\hat{\beta}_{i,T})$ exceeds the variance under constant betas by an additive term of size $h\sigma_f^2[(\beta_{i,0}^2 + \beta_{i,T}^2)/4 - (\beta_{i,0}\beta_{i,T})/2]/(T-T_0)\sigma_f^2$. Therefore, $Var(z_i)$ is larger due to the additional variation in both bias and variance of the estimator. Thus, the market risk premium in Equation (4.19) is necessarily more heavily attenuated than its constant beta counterpart in Equation (4.16).

As before, we have the t-statistic of the estimated market risk premium

$$\frac{\hat{t}_{\hat{\lambda}}}{\sqrt{N}} \rightarrow \sqrt{\Phi}\frac{\lambda_T}{\sqrt{Var(\hat{\lambda}_T) + (1-\Phi)\lambda_T^2}} < \frac{\lambda_T}{\sqrt{Var(\hat{\lambda}_T)}} \tag{4.20}$$

where again, $\Phi = Var(\beta_{i,T})/(Var(\beta_{i,T}) + Var(z_{i,T}))$. Using the same line of argumentation as for the attenuation bias of the point estimate, the t-statistic in Equation (4.20) will be subject to stronger downward-bias as in Equation (4.18).

It is likely that measurement error and beta are correlated. Hence, we now allow $\mathbb{E}(\beta_{i,T}z_{i,T}) \neq 0$. In particular, if $\beta_{i,T}$ is high, it is likely that $z_{i,T} < 0$, and the reverse. When the correlation between the estimation error and the level of beta is non-zero, the attenuation factor depends on the magnitude of that correlation. The market risk premium estimate converges in this situation to:

$$\hat{\lambda}_T \to \underbrace{\frac{Var(\beta_{i,T}) + Cov(\beta_{i,T}, z_{i,T})}{Var(\beta_{i,T}) + Var(z_{i,T}) + 2Cov(\beta_{i,T}, z_{i,T})}}_{=\Psi} \lambda_T \tag{4.21}$$

Taking the derivative of the attenuation factor Ψ with respect to $Corr(\beta_{i,T}, z_{i,T})$ we get

$$\frac{\partial \Psi}{\partial Corr(\beta_{i,T}, z_{i,T})} = \tag{4.22}$$

$$\Big(Var(z_{i,T}) - Var(\beta_{i,T})\Big) \frac{\sqrt{Var(\beta_{i,T})Var(z_{i,T})}}{\Big(Var(\beta_{i,T}) + Var(z_{i,T}) + 2Cov(\beta_{i,T}, z_{i,T})\Big)^2}$$

The sign of the derivative depends on the sign of $Var(z_{i,T}) - Var(\beta_{i,T})$. Therefore, introducing a negative correlation between the measurement error and beta will increase the attenuation bias if $Var(z_{i,T}) > Var(\beta_{i,T})$ and vice versa. Moreover, it turns out that a sufficiently negative correlation can actually cause a negative estimated risk-premium

$$\hat{\lambda}_T < 0 \iff Corr(\beta_{i,T}, z_{i,T}) < -\frac{Var(\beta_{i,T})}{Var(z_{i,T})} \tag{4.23}$$

provided that the true market risk premium is positive, i.e. $\lambda_T > 0$. This condition implies that the ordering of $\beta_{i,T}$ across assets is reversed, i.e. assets with low estimated betas $\hat{\beta}_{i,T}$ are always produced by assets with high returns and vice-versa.

Introducing correlation between the measurement error and beta leads to a more complicated effect on estimated variance of $\hat{\lambda}_T$ that now converges to

$$\begin{aligned}\hat{Var}(\hat{\lambda}_T) \to & \frac{Var(\beta_{T,i})}{Var(\beta_{i,T}) + Var(z_{i,T}) + 2Cov(\beta_{i,T}, z_{i,T})}(Var(\hat{\lambda}_T) + (1-\Psi)^2\lambda_T^2) + \\ & \frac{Var(z_{i,T})}{Var(\beta_{i,T}) + Var(z_{i,T}) + 2Cov(\beta_{i,T}, z_{i,T})}\Psi^2\lambda_T^2 + \\ & \frac{2Cov(\beta_{i,T}, z_{i,T})}{Var(\beta_{i,T}) + Var(z_{i,T}) + 2Cov(\beta_{i,T}, z_{i,T})}(1-\Psi)\Psi\lambda_T^2 \end{aligned} \tag{4.24}$$

Therefore, the estimated variance of $\hat{\lambda}_T$ is also clearly biased, but the direction of the bias is unclear. Unfortunately, this also extends to the corresponding t-statistic.

Price Staleness: $\beta_T = \beta_0 = \beta_1$ and $k > 0$

As in the last scenario we again assume that the betas are constant over time but observed returns suffer from stale prices. For the time being, assume that k is equal for all assets. In this case the multiplicative attenuation factor is equal for all assets and we can simply use $\hat{\beta}_{i,T} = (1-k/h)\beta_{i,T} + z_{i,T}$, with $z_{i,T} \sim (0, Var(\hat{\beta}_{i,T} - \beta_{i,T}))$.

In this situation, all betas are biased towards zero, and the absolute bias is greater for larger betas. Therefore, we again face a correlated measurement error problem. However, this time the correlation between measurement error and beta is known. Hence, if we choose the maximum possible length of the estimation window in the first stage, i.e. if $(T - T_0) \to \infty$ and therefore $Var(z_{i,T}) = 0$, we trivially have that

$$\hat{\lambda}_T \to (1-k/h)^{-1}\lambda_{T-k}. \tag{4.25}$$

That is, $\hat{\lambda}_T$ is actually inflated relative to the true λ_{T-k}, and therefore presumably to λ_T. Thus, unlike in the first stage, 'matching' in time the left and right-hand side does not lead to an attenuation bias but simply leads the estimated market risk premium to converge to the risk premium at the time of the last observed trade of the assets rather than the current time risk premium.

However, as long as the length of the estimation window $(T - T_0)$ is finite, there will be sampling error, and in addition to the inflation caused by downward biased betas, we will obtain attenuated estimates caused by the remaining zero-mean variation.

$$\hat{\lambda}_T \rightarrow \frac{(1-k/h)Var(\beta_{i,T})}{(1-k/h)^2 Var(\beta_{i,T}) + Var(z_{i,T})}\lambda_{T-k}$$

Note that there even exists a T_0 that balances the inflation and attenuation in such a way that $\hat{\lambda}_T \rightarrow \lambda_{T-k}$. This would be achieved if we could choose T_0 such that $Var(z_i) = ((k/h) - (k/h)^2)Var(\beta_{i,t})$. Even though it is not feasible to find a T_0 that balances the inflation and attenuation, it illustrates that in this scenario it is possible to obtain $\hat{\lambda}_T > 0$.

For the estimated variance of $\hat{Var}(\hat{\lambda}_T)$ under the current scenario we have

$$\hat{Var}(\hat{\lambda}_T) \rightarrow \tilde{\Phi}(1-k/h)^{-1}Var(\hat{\lambda}_T) + \tilde{\Phi}(1-\Phi)^2(1-k/h)^{-1}\lambda^2_{T-k} + (1-\tilde{\Phi})\tilde{\Phi}\lambda^2_{T-k} \tag{4.26}$$

where $\tilde{\Phi} = (1-k/h)Var(\beta_{i,T})/((1-k/h)^2 Var(\beta_{i,T}) + Var(z_{i,T}))$ and Φ is the same as before. Again, the variance of $\hat{\lambda}_T$ is biased but the direction of the bias is unclear. As before, this also extents to the corresponding t-statistic.

If we allow k to differ among assets, left and right-hand side are observation-wise matched, but not to each other. If k_i and $\beta_{i,T}$ are uncorrelated, we will simply converge to the limit in Equation (4.25) with k replaced with the average k_i and λ_{T-k} with the average λ_{T-k_i}. While increasing the level of aggregation of returns reduces bias, it also increases the variance of the beta estimator. In either case the estimated market risk premium is attenuated, and the estimator variance is biased.

In summary, we can conclude from our discussion that price measurement error, sampling error, and time-series variation in betas introduce variance-bias trade-offs into the first stage of the Fama-MacBeth procedure that affects the estimated market risk premium in the second stage in non-trivial ways. Both the point estimates and the estimator variance of the market risk premium are biased. Even in the most simple case without price measurement error or time-series variation in betas the estimated market risk premium and its t-statistic are already biased towards zero. This is simply due to sampling error since betas are always estimated from finite samples. If price measurement error and time-series variation in betas are introduced, biases of the point estimate and the estimator variance increase. We show that if betas are time varying and beta measurement error is negatively correlated with betas, then it is possible to obtain a negative point estimate of the market risk premium if the true premium is positive. The rather complicated structure of the biases of the estimated variance of the estimated market risk premium provides an indication why in previous studies standard methods of reducing estimated standard errors were unsuccessful (see, e.g., Grauer and Janmaat (2009)).

Our analytical results indicate a purely econometric rationale why previous studies that relied on standard OLS implementation of the Fama-MacBeth procedure found a flat market risk premium and why these studies lacked the power to reject the hypothesis that the market risk premium is zero. Additionally, we provide an econometric rationale for negative estimates of the market risk premium. However, our results also suggest that it is possible to estimate a positive significant market risk premium while only relying on a standard OLS implementations of the Fama-MacBeth procedure. We explore this possibility in the following sub-chapters.

4.3 Empirical Application

In this sub-chapter we empirically demonstrate the effect of price measurement error, sampling error, and time-series variation in betas on estimated betas and market risk premiums. We remain in the standard Fama and MacBeth (1973) framework with OLS estimates in the first and second stage estimation. However, we adapt the sample by forming a sub-sample based on stock liquidity and price staleness to create sub-samples of stocks that are most and least affected by price measurement error to subsequently show the properties of OLS estimates of betas and market risk premiums. We further vary sampling frequencies of returns and lengths of estimation windows of betas estimated from rolling-window OLS to demonstrate the effect on estimates of beta and market risk premiums. Since the effects of price measurement error, sampling error, and time-series variation in betas are interrelated this approach will not perfectly disentangle each of the effects from the others but will yield an intuition regarding the empirical importance of these statistical biases in widely used datasets.

4.3.1 Data

Our primary data sources are the CRSP database to obtain daily data and the TAQ database to obtain intra-day data for a subset of our sample. From the CRSP database we obtain daily total returns, end-of-day bid and ask prices, and daily trading volumes for all common stocks included in the CRSP database for all trading days between August 10, 1926 and December 31, 2013. Thus, our sample covers almost 90 years of stock market data encompassing different economic regimes and a time period with dramatically changing market structures and increasing trading activity. We also separately consider a more recent sub-period ranging from January 01, 1999 to December 31, 2013. In our analysis we aggregate data to daily, weekly, bi-weekly, and monthly figures. Rather than using calendar information we obtain weekly, bi-weekly, and monthly returns by aggregating the daily returns over 5, 10, and 20 trading days respectively.

We drop stocks from our sample that have less than 1,200 daily observations, i.e. 5 years, during the sample period. For the full sample period from August 10, 1926 to December 31, 2013 this yields 24,479 unique stocks. Since we have an unbalanced panel, we observe, on average, approximately 2,963 unique stocks on any given day. For the sub-period from January 01, 1999 to December 31, 2013 we have 11,499 unique stocks and, on average, approximately 4,813 stocks on any given day.

From the TAQ database we additionally obtain intra-day tick prices for a subset of stocks that were included in the S&P500 index between January 01, 1999 and December 31, 2013. Intra-day tick prices are aggregated to produce intra-day returns with 2, 3, 5, 6, 13, 26, 39, 78, or 390 observations per trading day. Assuming 6 1/2 trading hours per day, 390 observations correspond to one-minute returns.[9] We drop all observations of a given day if a stock has less than 130 return observations within that day. We observe for the S&P500 index universe in total 926 stocks and, on average, approximately 344 stocks on any given day. In all estimations we use the one month Treasury Bill rate as the risk-free rate and the return on the value-weighted CRSP market index as market portfolio.

[9]For example, the New York Stock Exchange opens from 09:30 a.m. to 04:00 p.m. (eastern time) and thus is open for 6 1/2 hours or 390 minutes per trading day.

4.3.2 Reducing Price Measurement Error

Price measurement error due to illiquidity and stale prices of stocks is a well known issue in empirical asset pricing. We measure the degree of price measurement error in the cross-section of stocks within each month along four dimensions: (1) the monthly $US trading volume, (2) the Amihud (2002) illiquidity measure, (3) the bid-ask spread relative to the mid price and (4) price staleness measured as the percentage of days within a month without trading activity.

While (4) reflects price staleness directly, (1) and (2) reflect that even if there is trading, observed prices can be noisy if trading is thin (see Asparouhova et al. (2010), Asparouhova et al. (2013)). However, there is a large amount of empirical literature establishing days without trading activity as a measure of illiquidity (see, e.g., Naes, Skjeltorp, and Odegaard (2011), Bekaert, Harvey, and Lundblad (2007), Lesmond (2005)). Furthermore, Bandi, Pirino, and Reno (2018b) and Bandi, Pirino, and Reno (2018a) show that price staleness is associated with a lack of trading volume or even low trading volumes. Note that Bandi et al. (2018b) distinguish between "staleness" and "idleness", where staleness describes a lack of price adjustment and idleness a lack of trading activity. According to this definition, prices could be stale even if there is some (low) trading volume. Therefore, our measure of price staleness as used in this chapter would be "idleness" in the sense of Bandi et al. (2018b), as we define "without trading activity" in (4) as zero trading volume. (3) reflects uncertainty about observed mid-prices that are used to obtain returns (see Blume and Stambaugh (1983)).

Bandi, Pirino, and Reno (2017) provide an economic rationale for the connection between illiquidity and price staleness. Market micro-structure founded theories of price formation with transaction costs, and asymmetric information suggests that informed traders react to new information not yet reflected in prices only if trading guarantees a profit net of transaction costs (see Hasbrouck and Ho (1987), Glosten and Milgrom (1985), Kyle (1985)). Thus, informed traders may choose not to trade in response to new information if transaction costs relative to fundamental values are too high. Additionally, Bandi et al. (2017) also argue that uninformed traders may not trade in a purely random way as they are sensitive to the absolute size of transaction costs, choosing not to trade if transaction costs are deemed to be too high. Since transaction costs are larger as assets become more illiquid, prices exhibit more staleness in illiquid samples.

We use all four measures jointly to create our illiquidity and price staleness sub-samples and estimate individual stock betas and, subsequently, market risk premiums in each sub-sample. In Equations (4.10) and (4.12) we show that price staleness represented by the parameter k potentially affects bias and standard errors of estimated betas. This is just another way of saying that low quality, noisy data leads to imprecise estimates if OLS is naïvely applied. Dropping low quality data from a broad sample of stocks, i.e. dropping stocks exhibiting illiquidity and price staleness, is equivalent to cleaning raw data as is commonly done in some way or another in the majority of empirical studies.

Summary statistics for all four measures for the sample period from August 10, 1926 to December 31, 2013 are reported in Table 4.3.1. For each measure, we report summary statistics for the total CRSP cross-section, the top two terciles, and the top tercile of the empirical distribution of the respective measure. The averages in Table 4.3.1 are averages in the cross-section and over time. Differences between the mean in the total cross-section and in the top terciles are rather large for all measures. Panel A in Table 4.3.1 shows that the average trading volume within the top tercile is three times the average trading volume in the total CRSP cross-section. Differences for the Amihud (2002) illiquidity measure in Panel B and bid-ask spreads in Panel D of the same table are even more pronounced. Table 4.3.1 Panel C shows the percentage of days in a month without trading activity. In the total CRSP cross-section the share of days without trading activity is 11.445%, which corresponds to approximately 2 days per month without any trading activity. The 95th percentile is at roughly 56.667%, which corresponds to approximately 11 days per month without any trading activity. Since the time period from 1926 to 2013 was subject to significant changes in market structures and dramatic increases in trading activity we also analyze a more recent sub-sample ranging from January 01, 1999 and December 31, 2013. Table 4.3.2 follows the same logic as Table 4.3.1 but uses only data for this more recent sub-sample. Since for this time period we also have intra-day data for stocks in the S&P500 index universe, we also report summary statistics for S&P500 index universe. The S&P500 index universe is an important benchmark as it is a commonly used alternative to the use of the total stock market and studied in numerous empirical surveys. While the S&P500 is comparable to the Top Tercile of the total CRSP sample in terms of liquidity, it remains an index intentionally constructed to be representative of the stock market. According to S&P

Table 4.3.1: Liquidity Measures for January 1926 - December 2013

Notes: The table shows liquidity measures for the time period January 1926 to December 2013. Panel A shows the monthly trading volume in million USD. Trading volume is constructed by multiplying the trading volume with the closing price, both variables are downloaded from CRSP. The monthly volume is found by summing over 20 trading days. Panel B shows the liquidity measure proposed in Amihud (2002). We divide the absolute value of the daily return by the trading volume in USD and report the average of 20 trading days for our monthly information. We disregard observations with zero trading volume and only take months with at least 10 observations into account. Returns and trading volumes are downloaded from CRSP. Panel C reports the average percentage of zero trading days during a period of 20 trading days. Panel D reports the average bid-ask spread as percentage of the price at the end of the trading day. Bid, ask, and end of day prices are downloaded from CRSP. The table shows averages over 20 trading days.

Sample	Mean	S. E.	5%	Median	95%
A: Average Trading Volume (in Mio. $)					
CRSP	125.917	1552.388	0.073	4.417	393.134
Top 2 Terciles	188.624	1898.175	1.818	15.068	630.545
Top Tercile	371.659	2672.058	17.181	66.459	1270.597
B: Amihud (2002) – Illiquidity measure					
CRSP	11.890	108.126	0.004	1.000	44.650
Top 2 Terciles	0.661	0.856	0.003	0.250	2.660
Top Tercile	0.064	0.068	0.001	0.036	0.212
C: Percentage of days without trading activity					
CRSP	11.445	18.900	0.000	2.143	56.667
Top 2 Terciles	1.410	2.106	0.000	0.237	6.406
Top Tercile	0.028	0.057	0.000	0.000	0.179
D: Bid-Ask spread in percent of price					
CRSP	4.360	6.806	0.181	2.433	14.234
Top 2 Terciles	1.555	1.120	0.145	1.354	3.621
Top Tercile	0.608	0.379	0.107	0.556	1.262

the index captures approximately 80% of the available market capitalization. Therefore, analyzing the S&P500 index universe should mitigate concerns regarding sample selection.

In comparison to the full sample period, the period from 1999 to 2013 is characterized by significantly higher liquidity and lower levels of price staleness in the entire cross-section and almost no remaining illiquidity or price staleness within the top terciles. This reflects the dramatic increase in trading activity in recent decades and indicates that strategies for testing asset pricing models should reflect market conditions during the sample period. The S&P500 index universe sub-sample is comparable to top tercile of the total CRSP cross-section in terms of the Amihud (2002) illiquidity and the price staleness measures. Bid-Ask spreads are slightly larger on average and show more variation than the CRSP top tercile but nonetheless are only a fraction of the average spreads in the total CRSP cross-section. The average trading volume in the S&P500 index universe is approximately 2.7 times the average trading volume in the top tercile of the total CRSP sample. These high trading volumes are likely an artifact of numerous mutual funds and ETFs tracking the S&P500 index.

Implications for the First Stage: Beta and Price Measurement Error

Traditionally researchers have attempted to reduce price measurement error in individual stock prices by sorting stocks into portfolios based on some characteristics of the data and then use the portfolio returns rather than the individual stock returns to test asset pricing theories. The idea behind this approach is that price measurement errors average out in the portfolios as prices for some stocks will be too high and prices of others will be too low due to mismeasurement (see, e.g., Blume (1970)). Therefore, the resulting portfolio betas should be estimated more precisely than the individual stock betas would have been. Since asset pricing models should hold for all assets, no matter whether they are portfolios or individual stocks, we would expect the same market risk premiums in either situation. However, forming portfolio creates its own set of statistical problems in the estimation of market risk premiums.

Lo and MacKinlay (1990) demonstrate analytically and through simulations that, in tests of asset pricing models, forming portfolios based on characteristics of the data creates potentially significant biases in test statistics even if the used characteristics are only

Table 4.3.2: Liquidity Measures for January 1999 - December 2013

The table shows liquidity measures for the time period January 1999 - December 2013. Panel A shows the monthly trading volume in million USD. Trading volume is constructed by multiplying the trading volume with the closing price, both variables are downloaded from CRSP. The monthly volume is found by summing over 20 trading days. Panel B shows the liquidity measure proposed in Amihud (2002). We divide the absolute value of the daily return by the trading volume in USD and report the average of 20 trading days for our monthly information. We disregard observations with zero trading volume and only take months with at least 10 observations into account. Returns and trading volumes are downloaded from CRSP. Panel C reports the average percentage of zero trading days during a period of 20 trading days. Panel D reports the average bid-ask spread as percentage of the price at the end of the trading day. Bid, ask, and end of day prices are downloaded from CRSP. The table shows averages over 20 trading days.

Sample	Mean	S. E.	5%	Median	95%
A: Average Trading Volume (in Mio. $)					
CRSP	307.340	2745.093	0.391	19.011	1153.552
Top 2 Terciles	459.960	3351.683	7.716	58.877	1682.080
Top Tercile	896.607	4699.796	66.933	226.054	3171.844
S&P500 Universe	2452.294	4304.520	157.535	1319.994	8138.240
B: Amihud (2002) – Illiquidity measure					
CRSP	7.435	90.224	0.001	0.134	31.049
Top 2 Terciles	0.105	0.160	0.000	0.026	0.500
Top Tercile	0.007	0.007	0.000	0.005	0.022
S&P500 Universe	0.007	0.050	0.000	0.001	0.015
C: Percentage of days without trading activity					
CRSP	5.391	12.723	0.000	0.000	34.256
Top 2 Terciles	0.064	0.165	0.000	0.000	0.476
Top Tercile	0.000	0.000	0.000	0.000	0.000
S&P500 Universe	0.002	0.024	0.000	0.000	0.000
D: Bid-Ask spread in percent of price					
CRSP	2.321	3.730	0.124	1.062	8.436
Top 2 Terciles	0.714	0.542	0.099	0.551	1.820
Top Tercile	0.279	0.139	0.071	0.266	0.516
S&P500 Universe	0.493	0.711	0.057	0.317	1.469

marginally correlated with statistics of interest. This raises concerns in particular regarding the common practice of using beta-sorted portfolios in tests of asset pricing models but also regarding size- or even industry-sorted portfolios. More recently, Ang, Liu, and Schwarz (2018) show analytically and empirically that forming portfolios in tests of asset pricing models leads to too large standard errors of the estimated market risk premiums. While forming portfolios does lead to a more precise estimation of the portfolio betas, especially in the presence of price measurement error, it also reduces the cross-sectional dispersion of betas that can be used to determine standard errors of the market risk premium. Since the standard errors of the market risk premium are largely determined by the cross-sectional distribution of betas, estimates of market risk premiums are less efficient if portfolios are used (see Ang et al. (2018)). If individual stock betas are used, then the cross section of betas is naturally more dispersed and therefore contains more information to estimate market risk premiums. Following this line of argumentation we use individual stock betas rather than portfolio betas in our analysis. The discussion in Ang et al. (2018) shows that it is crucially important for the precise estimation of market risk premiums to first precisely estimate betas of individual stocks. Rather than sorting individual stocks into portfolios we form a series of sub-samples of stocks that are decreasing in the average illiquidity and price staleness, while ensuring sufficient cross-sectional dispersion in betas.

We estimate two different rolling-window beta time-series for each individual stock in our total CRSP sample. First, we estimate betas at the end of each month using daily returns over rolling estimation windows of one year (240 trading days). Second, we estimate betas using monthly returns over rolling estimation windows of five years (60 months). Thus, we have two different time-series of monthly betas for each stock. Both choices of return frequency and estimation window length are common in the asset pricing literature. To find the stocks that are most and least affected by price measurement error we count, for each stock, the number of occurrences in the top tercile of trading volume, Amihud (2002) illiquidity, percentage of days without trading activity, and bid-ask spreads. The summary statistics for each of these measures are shown in Tables 4.3.1 and 4.3.2. Based on this count, we sort stocks into sub-samples of stocks that do not occur in any top tercile (”No Top” sample), stocks that occur in at least one top tercile (”Min. One Top” sample), stocks that occur in at least two top terciles (”Min. Two Top” sample), etc. until stocks that occur in all four top terciles (”Min. Four Top” sample). We expect stocks in the

”No Top” sample to be most affected by price measurement error and stocks in the ”Min. Four Top” sample to be least affected by price measurement error. Each of the sub-samples remains sufficiently large and dispersed for the subsequent estimation of market risk premiums. In the ”Min. One Top” sample we observe, on average, 1,487 stocks per month, in the ”Min. Two Top” sample we have, on average, 1,125 stocks per month; in the ”Min. three Top” sample, 872 stocks; and in the ”Min. Four Top” sample, 561 stocks.

In Table 4.3.3 we report for each of these sub-samples and for the total sample the average betas, the value of the min and max beta, and the average standard deviation of betas relative to the average standard deviation of betas in the total CRSP sample. The latter provides an indication for the reduction in the average standard deviation in the sub-samples as compared to the total CRSP cross-section.

Due to the panel structure of the data we can compute averages in two ways: either we compute cross-sectional averages in each month and then take the average over the resulting time-series of cross-sectional averages (”Cross-Section;Time”); or we compute the time-series average for each stock over its monthly betas and then take the cross-sectional average of all the stocks (”Time;Cross-Section”). If betas of individual stocks are unbiased we expect that it makes no difference how we choose to compute average betas and we expect an average beta close to one in either case. We expect the ”Cross-Section;Time” average betas to deviate from one if individual betas in the cross-section are biased such that the cross-sectional averages in every month are consistently under or over estimated. The ”Time;Cross-Section” average betas would deviate from one if individual betas for some stocks are consistently under or over estimated over time. Panel A of Table 4.3.3 shows ”Cross-Section;Time” and ”Time;Cross-Section” average betas computed from betas that are estimated over rolling windows of one year of daily returns. Panel B of the same table shows the averages for betas estimated over rolling windows of five years of monthly returns.

For Panel A of Table 4.3.3 we find a rather large difference between ”Cross-Section;Time” average beta and ”Time;Cross-Section” average beta only in the total CRSP sample and the ”No Top” sample. The average beta in the ”No Top” sample is 0.825 and 0.499 for the ”Cross-Section;Time” average beta and ”Time;Cross-Section” average beta respectively. Thus, betas of stocks that are more severely affected by illiquidity and price staleness tend to be biased downward. This is consistent with Equation (4.10) and with results reported in Scholes and Williams (1977). For a fixed level of return aggregation and a fixed estimation window length we expect betas of individual stocks to be biased downward if there is price measurement error. For the same degree of price measurement error, we expect that the bias decreases as we increase return aggregation, i.e. use monthly rather than daily returns to estimate betas. Indeed, for Panel B we find a ”Time;Cross-Section” average beta close to one for the ”No Top” sample.

The columns “REL STD β” in Table 4.3.3 show the percentages of the average standard deviation of the estimated beta relative to the average in the total CRSP sample. Thus, we report the decrease (“REL STD β”<1) or increase (“REL STD β”>1) in the average standard deviation of beta relative to the average standard deviation of beta in total CRSP sample. For Panels A and B, ”Cross-Section;Time” averages, and ”Time;Cross-Section” averages we observe the same pattern. Consistent with Equation (4.12) the average standard deviation in betas is increased relative to the CRSP average in the ”No Top” sample but decreases as the level of liquidity and price staleness decreases, i.e. as we move from ”Min. Two Top” to ”Min. Four Top”. The average standard deviation decreases slightly stronger in Panel A than Panel B. The decreasing standard deviations of betas in the more liquid samples provides an interesting intuition regarding the beta estimator suggested in Cosemans et al. (2016) that yields a positive and significant market risk premium. They obtain betas from time-series of stock returns using OLS and shrink them towards prior betas obtained from valuation information. The shrinkage weights depend on the relative standard deviation such that the weight attached to the OLS estimate of beta is larger the lower the standard deviation of the beta. Thus, if applied to the total CRSP cross-section this estimator would attach larger weights to more liquid stocks as their OLS betas are more informative.

We report average betas and corresponding relative average standard deviations for the sample period from January 1999 to December 2013 in Table 4.3.4. The results for this more recent period are consistent with our findings for the full sample period. In particular we find that the S&P500 index universe sample is comparable to the most liquid sub-sample of the CRSP cross-section. The average betas are consistently close to one in both Panels and for “Cross-Section; Time” and “Time; Cross-Section” averages suggests low biases in individual stock betas. The average standard deviation in betas in the S&P500 index universe sample varies between approximately 32% - 60% of the average standard

Table 4.3.3: Properties of Beta estimations for January 1926 - December 2013

The table shows properties of beta estimations on several samples formed on liquidity consideration. The sample ranges from January 1926 to December 2013. Panel A shows the descriptives for the estimators using daily returns. Panel B shows the descriptives for the estimators using monthly returns. The rows titled "Cross-Section; Time" report the statistics constructed by taking the average over the stocks for each months first and taking the average over time next. The rows titled "Time; Cross-Section" report statistics constructed by taking the average over time for each cross-section first and taking the average over the cross-sections next. The rows denoted by "CRSP" shows the statistics for the full CRSP sample. The rows denoted by "No Top" denote the statistics for the sample that does only contain stocks that are never in the top tercile of liquidity measures. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly. The column "REL STD β" shows the percentages of the average standard deviation of the estimated β relative to the average in the total CRSP sample. For example, the value of 1.414 in the row "No Top" of the Panel A using the method "Cross-Section; Time" means that the average standard deviation is 1.414 times as large as the average standard deviation of the total CRSP sample.

	A: 240 Days				B: 60 Months			
	AVG β	MIN β	MAX β	STD β	AVG β	MIN β	MAX β	STD β
Cross-Section; Time								
CRSP	0.948	-11.101	8.145	1.000	1.165	-3.996	6.848	1.000
No Top	0.825	-11.101	8.145	1.414	1.154	-3.996	6.562	1.281
Min. One Top	1.020	-3.977	7.335	0.556	1.143	-3.743	6.848	0.772
Min. Two Top	1.010	-3.977	6.217	0.443	1.112	-3.743	6.848	0.682
Min. Three Top	1.007	-3.744	5.478	0.390	1.088	-3.743	6.848	0.637
Min. Four Top	1.070	-2.009	5.478	0.350	1.099	-1.616	6.848	0.600
Time; Cross-Section								
CRSP	0.710	-11.101	8.145	1.000	1.123	-3.996	6.848	1.000
No Top	0.499	-11.101	8.145	1.340	0.963	-3.996	6.562	1.112
Min. One Top	0.983	-3.977	7.335	0.563	1.299	-3.743	6.848	0.876
Min. Two Top	1.055	-3.977	6.217	0.451	1.319	-3.743	6.848	0.807
Min. Three Top	1.101	-3.744	5.478	0.396	1.337	-3.743	6.848	0.772
Min. Four Top	1.160	-2.009	5.478	0.319	1.366	-1.616	6.848	0.690

deviation in betas in the total CRSP sample.

Implications for the Second Stage: Market Risk Premium and Price Measurement Error

Since our goal is the estimation of market risk premiums we now investigate the impact of price measurement error on the second stage of Fama-MacBeth procedure. We use the same estimated betas as in Tables 4.3.3 and 4.3.4 to estimate the market risk premium in each sub-sample. We report monthly and annualized market risk premiums and the corresponding t-statistics and p-values in Tables 4.3.5 and 4.3.6. Panels and rows of Tables 4.3.5 and 4.3.6 follow the same logic as Tables 4.3.3 and 4.3.4. The reported p-values correspond to a one-sided t-test of the null hypothesis that the market risk premium $\lambda \leq 0$ against the alternative hypothesis $\lambda > 0$. Using a one-sided rather than a two-sided t-test corresponds more closely with the purpose of our analysis since our goal is to demonstrate that the market risk premium implied by the CAPM is positive after we control for statistical biases.

It is difficult to determine a "correct" magnitude for the market risk premium. Theoretically the market risk premium is the expected return on the market portfolio in excess of the risk-free rate. The annualized average return of the value-weighted CRSP market index in excess of the one month Treasury Bill rate for the period from August 10, 1926 to December 31, 2013 was 6.970% using the arithmetic mean and 5.511% using the geometric mean. For the period from January 01, 1999 to December 31, 2013 it was 5.318% and 3.185% for the arithmetic and geometric mean respectively. These numbers provide some indication for a sensible range of the market risk premium.

In Panel A of Table 4.3.5, using rolling windows of 240 daily returns to estimate betas, we find a negative insignificant market risk premium similar to results reported in earlier studies. For the "Min. One Top" sample market the estimated risk premium is positive but close to zero and insignificant. However, for the "Min. Two Top" sample to the "Min. Four Top" sample, we find market risk premiums that increase in magnitude and significance. The market risk premium estimated in the "Min. Four Top" sample is 3.926% per annum with a t-statistic of 2.114 and is therefore significant at all conventional levels.[10] In Panel B

[10]Given a t-statistics of roughly 2.1 the market risk premium would also be considered statistically signif-

Table 4.3.4: Properties of Beta estimations for January 1999 - December 2013

The table shows properties of beta estimations on several samples formed on liquidity consideration. The sample ranges from January 1999 to December 2013. Panel A shows the descriptives for the estimators using daily returns. Panel B shows the descriptives for the estimators using monthly returns. The rows titled "Cross-Section; Time" report the statistics constructed by taking the average over the stocks for each months first and taking the average over time next. The rows titled "Time; Cross-Section" report statistics constructed by taking the average over time for each cross-section first and taking the average over the cross-sections next. The rows denoted by "CRSP" shows the statistics for the total CRSP sample. The rows denoted by "No Top" denote the statistics for the sample that does only contain stocks that are never in the top tercile of liquidity measures. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly. The column "STD β" shows the percentages of the average standard deviation of the estimated β relative to the average in the total CRSP sample. For example, the value of 1.402 in the row "No Top" of the Panel A using the method "Cross-Section; Time" means that the average standard deviation is 1.402 times as large as the average standard deviation of the total CRSP sample.

	A: 240 Days				**B: 60 Months**			
	AVG β	MIN β	MAX β	STD β	AVG β	MIN β	MAX β	STD β
Cross-Section; Time								
CRSP	0.892	-3.279	5.478	1.000	1.097	-2.372	6.848	1.000
No Top	0.517	-3.279	4.300	1.402	0.902	-2.372	6.562	1.198
Min. One Top	1.120	-2.359	5.478	0.752	1.206	-1.277	6.848	0.888
Min. Two Top	1.150	-1.563	5.478	0.578	1.177	-0.821	6.848	0.742
Min. Three Top	1.143	-1.563	5.478	0.494	1.154	-0.821	5.527	0.668
Min. Four Top	1.140	-0.667	5.478	0.454	1.143	-0.821	5.527	0.627
S&P 500	1.063	-0.667	5.478	0.328	1.062	-0.562	5.309	0.487
Time; Cross-Section								
CRSP	0.789	-3.279	5.478	1.000	1.114	-2.372	6.848	1.000
No Top	0.430	-3.279	4.300	1.300	0.879	-2.372	6.562	1.122
Min. One Top	1.054	-2.359	5.478	0.777	1.292	-1.277	6.848	0.908
Min. Two Top	1.140	-1.563	5.478	0.523	1.254	-0.821	6.848	0.673
Min. Three Top	1.143	-1.563	5.478	0.430	1.216	-0.821	5.527	0.581
Min. Four Top	1.180	-0.667	5.478	0.393	1.234	-0.821	5.527	0.532
S&P 500	1.021	-0.667	5.478	0.246	1.052	-0.562	5.309	0.398

of the same table, using rolling windows of 60 monthly returns to estimate betas, we find consistently positive significant market risk premiums in the total CRSP sample and in the "Min. Two Top" sample to the "Min. Four Top" sample, albeit again with increasing magnitude of the point estimates and in increasing significance. For the total CRSP cross-section we find a market risk premium of 2.678% per annum with a t-statistic of 1.569. The t-statistics in the "Min. One Top" sample to the "Min. Four Top" samples are between approximately 2.6 and 3.9, while point estimates lie between approximately 4.628% and 8.271% per annum.[11]

Table 4.3.5: Market Risk Premiums January 1926 to December 2013

The table shows the estimated risk premium for the sample period January 1926 to December 2013. The estimation was done by using the method proposed in Fama and MacBeth (1973). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A uses daily returns of a time period of 240 days and Panel B uses monthly returns of a time period of 60 months. In the second step these betas are regressed on the realized return of single stocks over the immediately following 20 days. The column denoted by "MRP (M)" shows the monthly estimator, the column denoted by "MRP (Y)" shows the annualized risk premium, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $\lambda > 0$. The rows denoted by "CRSP" shows the statistics for the total CRSP sample. The rows denoted by "No Top" denote the statistics for the sample that only contains stocks that are never in the top tercile of liquidity measures. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly.

	A: 240 Days				B: 60 Months			
	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val
CRSP Total	-0.132	-1.569	-1.199	0.885	0.221	2.678	1.569	0.058
Min. One Top	-0.025	-0.299	-0.200	0.579	0.378	4.628	2.578	0.005
Min. Two Top	0.117	1.407	0.856	0.196	0.394	4.834	2.670	0.004
Min. Three Top	0.175	2.121	1.283	0.100	0.464	5.711	3.038	0.001
Min. Four Top	0.321	3.926	2.114	0.017	0.664	8.271	3.901	0.000

We also report results for the period from January 01, 1999 to December 31, 2013 in Table 4.3.6. We observe a similar pattern of market risk premiums across the two panels and across the different sub-samples; however, the significance is consistently lower. For

icant at the 5% level even if the premise of using a one-sided rather than two-sided test would be rejected.

[11]Again the t-statistics for the "Min. One Top" sample to the "Min. Four Top" imply statistical significance at the 1% level even if the premise of a one-sided test would be rejected.

Panel A, none of the market risk premiums is statistically significant and obtains negative point estimates for the total CRSP sample, the "Min. One Top" and "Min. Two Top" sub-samples. Only the "Min. Three Top", "Min. Four Top" and "S&P 500" samples yield positive point estimates in a sensible range given the historic average return of the market portfolio during the sample period. However, the significance is substantially lower. For Panel B, point estimates of market risk premium are higher in magnitude compared to Panel A and exhibit larger t-statistics.

Table 4.3.6: Market Risk Premiums January 1999 to December 2013

The table shows the estimated risk premium for the sample period January 1963 to December 1991. The estimation was done by using the method proposed in Fama and MacBeth (1973). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A uses daily returns of a time period of 240 days and Panel B uses monthly returns of a time period of 60 months. In the second step these betas are regressed on the realized return of single stocks over the immediately following 20 days. The column denoted by "MRP (M)" shows the monthly estimator, the column denoted by "MRP (Y)" shows the annualized risk premium, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $\lambda > 0$. The rows denoted by "CRSP" shows the statistics for the total CRSP sample. The rows denoted by "No Top" denote the statistics for the sample that does only contain stocks that are never in the top tercile of liquidity measures. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly.

	A: 240 Days				**B: 60 Months**			
	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val
CRSP Total	-0.185	-2.193	-0.568	0.714	0.350	4.285	1.199	0.116
Min. One Top	-0.192	-2.276	-0.498	0.690	0.401	4.921	1.206	0.115
Min. Two Top	-0.012	-0.144	-0.028	0.511	0.508	6.266	1.498	0.068
Min. Three Top	0.255	3.106	0.562	0.287	0.533	6.584	1.612	0.054
Min. Four Top	0.438	5.381	0.946	0.173	0.571	7.077	1.703	0.045
S&P 500	0.504	6.213	1.104	0.136	0.442	5.437	1.291	0.099

Overall, the effect of reducing price measurement error by reducing illiquidity and price staleness in the estimation sample appears to be larger in the long-run sample than in the more recent sample. In the long-run sample from 1926 to 2013 we already obtained positive and highly significant market risk premiums by simply disregarding illiquid stocks from the sample. Point estimates in the S&P500 universe are comparable to the estimates in the “Min. Four Top” sub-samples in both panels. Generally, we obtain point estimates of the market risk premium for the “Min. Three Top” and the “Min. Four Top” sub-samples that consistent across Panels A and B in both Table 4.3.5 and Table 4.3.6. Only the t-statistics are substantially smaller in the more recent sample period.

4.3.3 Reducing Sampling Error

The choice of frequency of returns, used in both the estimation of betas and in the test of asset pricing models, is all too often driven by convenience. Arguably the most popular choices are monthly or daily returns, although weekly, quarterly or even yearly returns are also sometimes used. More recently the use of intra-day returns has gained popularity, but applications focus mostly on forecasting variances at the individual stock level rather than testing asset pricing models. In the estimation of betas, for a given length of an estimation window the number of observations used by the estimator depends on the sampling frequency of returns. Everything else being equal, the lower the chosen return frequency the higher the sampling error in estimated betas. In the following analysis we focus on the S&P500 index universe sub-sample and the time period from January 01, 1999 to December 31, 2013. This sub-sample is comparable to the “Min. Four Top” sub-sample but allows to increase the sample frequency of returns to intra-day observations. Thus, focusing on a sample with high liquidity and low levels of price staleness helps to disentangle the effect from price measurement error from the effects of sampling error and time-series variation. Additionally, the S&P500 index is, by construction, a representative market index covering most of the available market capitalization and therefore, there is less concern regarding sample selection.

Implications for the First Stage: Return Frequency & Beta

We assess the influence of the frequency of returns on the beta estimator by repeatedly estimating betas on the same rolling-windows but successively decreasing the sampling frequency from intra-day returns to monthly returns within each window. This is equivalent to stepwise increases in the aggregation parameter h in sub-chapter 4.2, where each decrease in return frequency corresponds with an increase in the level of return aggregation

h. Intra-day returns used in this sub-chapter refer to two return observation per day.[12]

We analyze the effect of different return frequencies by estimating for each stock in the S&P500 index universe sample at the end of each month its beta over a rolling estimation window of either 12 months or 60 months but vary the return frequencies within each of the estimation windows. We report summary statistics for these beta estimations in Table 4.3.7. The table follows the same logic as table 4.3.4. However, the rows refer now to the different return frequencies ranging from intra-day to monthly returns. Panel A shows results for the 12 months estimation window, while panel be shows results for the 60 months estimation window. As before, in the upper half of Table 4.3.7 betas are first averaged cross-sectionally in each month, then the time-series average over the monthly averages is computed. The lower half shows results if time-series averages are computed for each stock, then the cross-sectional average over those averages is computed. Since betas estimated on intra-day returns have the lowest average standard deviation we report all average beta standard deviations relative to the average standard deviation of the betas estimated from intra-day returns.

For both panels of Table 4.3.7 average betas are close to one for "Cross-Section; Time" and "Time; Cross-Section" averages. Thus, individual stock betas appear to be on average unbiased. This is not surprising as the different sampling frequencies of returns should only affect the standard deviation of betas but not point estimates. That the average standard deviation of betas increases as the frequency of returns decreases, i.e. as the number of observations in the estimation window decreases, is almost tautological. However, the magnitude of the increase is noteworthy. For example in Panel A of Table 4.3.7 aggregating returns from intra-day to daily returns within an estimation window reduces the sample size by the factor $1/h = 1/2$, i.e. from 480 observation within 12 months for intra-day returns to 240 observation for daily returns. This reduction in sample size leads to a roughly 7 times increase in average standard deviations. Further decreasing the return frequency to weekly returns reduces the sample size by the factor $1/10$ to 48 observation whereas the average standard deviation is roughly 40 times as large as for the intra-day returns. Panel B of Table 4.3.7 shows that increasing the length of the estimation window to some extent mitigates

[12] In the next sub-chapter we report results on various choices of intra-day return frequencies. The admittedly unusual choice of intra-day return frequency of two observations per day is motived by our finding that this choices leads to the highest t-statistics when short estimation windows are used. We discuss a rational for this choice and report results for other choices of intra-day return frequencies in the next sub-chapter.

the effect of reducing the return frequency as increases in the average standard deviations are less pronounced. This may be due to time-series variation in betas.

Implications for the Second Stage: Return Frequency & Market Risk Premium

Noisy estimates of beta lead estimates of the market risk premiums to be biased towards zero. Thus, if betas are estimated with greater average standard deviations, then estimates of the market risk premium are more heavily attenuated. Our analytical discussion indicates that this effect is further exaggerated if betas are also time-varying and if estimated betas and measurement errors are correlated. Since we have seen that decreasing sampling frequencies of returns lead to substantial increases in the average standard deviation of estimated betas, we expect to find lower point estimates of the market risk premium for betas estimated using lower return frequency data. Noisy estimates of beta also lead to biased statistical inference. The t-statistics are clearly biased towards zero if estimated betas are just noisy but constant over time and estimated betas and measurement errors are correlated. However, in case of time-varying betas and if estimated betas and measurement errors are correlated the direction of the bias cannot easily be determined.

We use the estimated betas summarized in table 4.3.7 to estimate market risk premiums. Thus, for each length of the estimation window and for each of the return frequency we obtain one estimate for the market risk premium. The results of the estimation are reported in Table 4.3.8. The table follows the same logic as the tables in the previous section, whereas now the rows refer to the different return frequencies. In Panel A of Table 4.3.8 the point estimate exhibit the predicted pattern of increasingly attenuated market risk premium estimates except for the intra-day returns. Point estimates for the market risk premium decease as the return frequency decreases from daily to monthly and therefore the average standard deviation of estimated betas increases. Increasing the return frequency to intra-day returns leads to a lower point estimate.

In Panel B of Table 4.3.8 we observe the opposite pattern as point estimates increase from a negative market risk premium for intra-day returns to a point estimate of almost 5% for monthly returns. In both panels of Table 4.3.8 estimated market risk premiums are statistically insignificant for all return frequencies. These results hint towards a more complex interconnection between the length of the estimation window and the return

Table 4.3.7: Market Risk Premiums January 1999 to December 2013

The table shows properties of beta estimations in the S&P500 sample using different return frequencies ranging from high-frequency to monthly returns. The sample ranges from January 1999 to December 2013. Panel A shows the descriptives for the estimators using a 12 months rolling estimation window. Panel B shows the descriptives for the estimators using a 60 months rolling estimation window. The rows titled "Cross-Section; Time" report the statistics constructed by taking the average over the stocks for each months first and taking the average over time next. The rows titled "Time; Cross-Section" report statistics constructed by taking the average over time for each cross-section first and taking the average over the cross-sections next. The rows denoted by "Intra-Day" shows the statistics if beta estimates are based on intra-day returns using two observations per day. The rows denoted by "Daily", "Weekly", "Bi-Weekly", and "Monthly" denote the statistics if beta estimates are based on daily, weekly, bi-weekly, or monthly returns respectively. The column "STD β" shows the percentages of the average standard deviation of the estimated β relative to the average under the use of high-frequency returns. For example, the value of 6.929 in the row "Daily" in Panel A using the method "Cross-Section; Time" means that the average standard deviation is 6.929 times as large as the average standard deviation if intra-day returns were used.

	A: 12 Months				**B: 60 Months**			
	AVG β	MIN β	MAX β	STD β	AVG β	MIN β	MAX β	STD β
Cross-Section; Time								
Intra-Day	1.065	-0.765	4.108	1.000	1.083	-0.351	2.931	1.000
Daily	1.059	-1.325	5.478	6.929	1.021	-0.394	3.427	5.100
Weekly	1.084	-1.766	6.732	37.836	1.054	-0.373	4.309	25.618
Bi-Weekly	1.088	-2.891	7.847	89.949	1.068	-0.705	4.536	57.862
Monthly	1.090	-5.661	9.239	186.986	1.059	-0.562	5.309	117.035
Time; Cross-Section								
Intra-Day	1.065	-0.765	4.108	1.000	1.083	-0.351	2.931	1.000
Daily	1.020	-1.325	5.478	7.565	1.010	-0.394	3.427	5.452
Weekly	1.041	-1.766	6.732	41.249	1.042	-0.373	4.309	27.079
Bi-Weekly	1.054	-2.891	7.847	94.750	1.059	-0.705	4.536	59.722
Monthly	1.042	-5.661	9.239	203.190	1.049	-0.562	5.309	124.347

frequency and thus can indicate time-series variation in betas. Time-series variation in betas is well documented in the literature (see Jagannathan and Wang (1996), Fabozzi and Francis (1978), Rosenberg and Guy (1976), Blume (1975)). If betas were constant during the sample period the only difference between Panel A and B within the same row would be sample size. Since Panel B betas are based on a five times larger sample size for the same return frequency the standard deviation of betas in Panel B should be significantly lower and estimated market risk premiums less heavily attenuated. Thus, we would expect the market risk premiums in Panels A and B to move in the same rather than opposing directions as return frequency decreases.

Table 4.3.8: Measurement MRP

The table shows the estimated risk premium for the sample period January 1999 to December 2013. The estimation was done by using the method proposed in Fama and MacBeth (1973). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A uses daily returns of a time period of 240 days and Panel B uses monthly returns of a time period of 60 months. In the second step these betas are regressed on the realized return of single stocks over the immediately following 20 days. The column denoted by "MRP (M)" shows the monthly estimator, the column denoted by "MRP (Y)" shows the annualized risk premium, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $\lambda > 0$. The rows denoted by "Intra-Day" shows the statistics if beta estimates are based on intra-day returns using two observations per day. The rows denoted by "Daily", "Weekly", "Bi-Weekly", and "Monthly" denote the statistics if beta estimates are based on daily, weekly, bi-weekly, or monthly returns respectively.

	A: 12 Months				**B: 60 Months**			
	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val
Intra-Day	0.316	3.856	0.615	0.270	-0.080	-0.956	-0.154	0.561
Daily	0.469	5.779	1.047	0.148	0.226	2.750	0.473	0.318
Weekly	0.247	3.008	0.712	0.239	0.286	3.490	0.661	0.255
Bi-Weekly	0.217	2.630	0.772	0.221	0.392	4.812	0.975	0.165
Monthly	0.235	2.857	1.085	0.140	0.403	4.949	1.188	0.118

4.3.4 Time-Series Variation

The results of the preceding sub-chapter indicate a more complex relation between return frequency and estimation window length that could be explained by the effect of time-series variation in betas. Therefore, we further investigate the connection between return frequency and estimation window length in the following sub-chapters. For the same reasons as discussed before we remain in the S&P500 index universe sample and the within the sample period from January 01, 1999 to December 31, 2013.

Implications on the First Stage: Estimation Windows & Beta

We estimate rolling window betas using estimation window lengths of 20, 60, 120, 240, and 1200 days. Within each estimation window length we use intra-day, daily and monthly returns to estimate betas. Hence, we have 15 different time-series of estimated betas referring to the different return frequency and estimation window lengths for each individual stock in the sample. Summary statistics for each combination of return frequency and estimation window length are reported in Table 4.3.9. The table again follows the same logic as the corresponding preceding tables, whereas now rows denote different estimation window lengths and panels denote different return sampling frequencies within each estimation window. Panel A shows the summary statistics for intra-day returns. Panels B and C show summary statistics for daily returns and monthly returns respectively. We report average standard deviations relative to the average standard deviation of the 20 days estimation window for each return frequency.

The results essentially indicate the same effect as documented in sub-chapter 4.3.3. Increasing the sample size, i.e. the number of return observation within the estimation window, leads to a lower average standard deviation in betas. We have already established that keeping the estimation window length constant when increasing the return frequency leads to a decrease in the average standard deviation. Table 4.3.9 shows that obviously the same can be achieved by keeping the return frequency constant and increasing estimation window. Both leads to an increase in the number of data points used in the estimation and thus naturally to a lower standard deviation.

Implications on the Second Stage: Estimation Windows & Market Risk Premiums

To analyze the effect of the choice of the estimation window length we use the estimated betas summarized in table 4.3.9 and estimate market risk premiums. The results are shown

Table 4.3.9: Market Risk Premiums

The table shows properties of beta estimations in the S&P500 sample using different estimation window lengths. The sample ranges from January 1999 to December 2013. Panel A shows the descriptives for the estimators using high-frequency returns. Panel B shows the descriptives for the estimators using daily returns, and Panel C shows the descriptives for the estimator using monthly returns. The rows titled "Cross-Section; Time" report the statistics constructed by taking the average over the stocks for each months first and taking the average over time next. The rows titled "Time; Cross-Section" report statistics constructed by taking the average over time for each cross-section first and taking the average over the cross-sections next. The rows indicate the length of the rolling estimation window used in the estimation of betas. In Panel C one month corresponds to 20 days, thus for example the row "1200 Days" uses 60 monthly return observations. The column "STD β" shows the percentages of the average standard deviation of the estimated β relative to the average under the use of a 20 days estimation window. For example, the value of 29.638 in the row "60 Days" in Panel A using the method "Cross-Section; Time" means that the average standard deviation is roughly 2/3 smaller as the average standard deviation compared to the 20 days estimation window.

	A: High Freq.				**B: Daily**				**C: Monthly**			
	AVG β	MIN β	MAX β	STD β	AVG β	MIN β	MAX β	STD β	AVG β	MIN β	MAX β	STD β
						Cross-Section; Time						
20 Days	1.054	-4.157	7.805	1.000	1.063	-7.464	20.519	1.000	–	–	–	–
60 Days	1.055	-2.471	5.114	0.317	1.061	-2.153	9.032	0.280	1.159	-122.722	138.382	1.000
120 Days	1.059	-1.409	4.610	0.155	1.062	-1.106	7.780	0.130	1.111	-17.628	22.679	0.091
240 Days	1.065	-0.765	4.108	0.080	1.063	-0.667	5.478	0.061	1.092	-5.661	9.239	0.030
1200 Days	1.083	-0.351	2.931	0.021	1.025	-0.284	3.427	0.011	1.062	-0.562	5.309	0.005
						Time; Cross-Section						
20 Days	1.049	-4.157	7.805	1.000	1.003	-7.464	20.519	1.000	–	–	–	–
60 Days	1.052	-2.471	5.114	0.318	1.004	-2.153	9.032	0.281	1.109	-122.722	138.382	1.000
120 Days	1.057	-1.409	4.610	0.156	1.011	-1.106	7.780	0.129	1.055	-17.628	22.679	0.103
240 Days	1.065	-0.765	4.108	0.081	1.021	-0.667	5.478	0.059	1.039	-5.661	9.239	0.033
1200 Days	1.083	-0.351	2.931	0.021	1.015	-0.284	3.427	0.011	1.052	-0.562	5.309	0.005

in table 4.3.10. The table follows the same logic as the corresponding tables in the previous sub-chapter, whereas the rows indicate the choice of the estimation window length used in the estimation of betas and Panels A to C indicate intra-day, daily, and monthly returns within each estimation window respectively.

In line with the result in Table 4.3.8 the estimated market risk premium in Panel A of 4.3.10, for betas estimated over 1200 days of intra-day returns, is negative and insignificant. However, for shorter estimation windows we obtain point estimates of the market risk premium and corresponding t-statistics that are both substantially higher. We find a similar pattern in Panel B of 4.3.10 for daily returns. For the intra-day return estimates in Panel A and the daily returns estimates in Panel B we estimate market risk premiums between approximately 5% and 8.5% per annum. In Panel A risk premiums based on betas estimated over up to 120 days of intra-day returns and in Panel B risk premiums based on betas estimated over 20 days of daily returns are also statistically significant at the 10% level for the one-sided test. Market risk premiums in Panel A for a estimation window length of 1200 days are negative. Statistical significance improves for shorter estimation windows and for higher sampling frequencies. This is consistent with our theoretical discussion. Shorter estimation windows lead to lower risk of biased beta estimates if betas are time-varying and thus to lower measurement error in the second stage regression. Moreover, increasing the sampling frequency for a particular estimation window effectively increases the sample size and thus leads to a lower variance of betas which again helps mitigating attenuation bias in the second stage regression.

So far we have simply used two return observations per day to describe intra-day returns. However, intra-day returns can be observed at much higher frequencies. Increasing return frequencies comes at the cost of significantly exaggerating issues of market micro-structure noise and non-synchronicity between the return process of the individual stock returns and the returns process of the market portfolio. The higher the return frequency the greater the chance that some stocks exhibit stale prices during a day. Covariances are affected to a greater degree than variances and the optimal sampling frequency in light of market micro-structure noise and non-synchronicity can be substantially larger for covariance than variances (see Epps (1979), Reno (2003), Hayashi and Yoshida (2005), Boudt, Laurent, Lunde, Quaedvlieg, and Sauri (2017)). Therefore, intra-day return frequencies cannot be increased without bound as at some point price measurement error induced by market

Table 4.3.10: Market Risk Premiums

The table shows shows the estimated risk premium for the sample period January 1999 to December 2013. The estimation was done by using the method proposed in Fama and MacBeth (1973). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A shows the descriptives for the estimators using high-frequency returns. Panel B shows the descriptives for the estimators using daily returns, and Panel C shows the descriptives for the estimator using monthly returns. The column denoted by "MRP (M)" shows the monthly estimator, the column denoted by "MRP (Y)" shows the annualized risk premium, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $\lambda > 0$. The rows indicate the length of the rolling estimation window used in the estimation of betas in the first stage. In Panel C one month corresponds to 20 days, thus for example the row "1200 Days" uses 60 monthly return observations.

	A: Intra-Day				**B: Daily**				**C: Monthly**			
	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val
20 Days	0.539	6.662	1.596	0.056	0.422	5.183	1.536	0.063	0.000	0.000	0.000	0.000
60 Days	0.685	8.533	1.569	0.059	0.421	5.172	1.146	0.127	0.081	0.975	0.929	0.177
120 Days	0.677	8.435	1.396	0.082	0.495	6.110	1.197	0.116	0.219	2.659	1.512	0.066
240 Days	0.316	3.856	0.615	0.270	0.469	5.779	1.047	0.148	0.235	2.857	1.085	0.140
1200 Days	-0.080	-0.956	-0.154	0.561	0.226	2.750	0.473	0.318	0.403	4.949	1.188	0.118

micro structure noise and non-synchronicity will outweigh the benefits form an increased sample size. Thus, the best choice of intra-day return frequency becomes an empirical question. In Table 4.3.11 Panel A to I we report estimated market risk premiums using betas estimated over estimation windows 20 to 12000 days of intra-day returns with two observations per day (Panel A) to 390 observations per day (Panel I). Point estimates of the market risk premium are rather stable for most choices intra-day return frequency, except for Panel I. However, consistent with the argument that too return frequency lead to noisier beta estimates we find that t-statistics decrease as the intra-day sampling frequency increases and t-statistics are highest for two return observations per day.

Table 4.3.11: Market Risk Premiums Intra-Day Return

The table shows shows the estimated risk premium for the sample period January 1999 to December 2013. The estimation was done by using the method proposed in Fama and MacBeth (1973). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A shows the descriptives for the estimators using high-frequency returns. Panel B shows the descriptives for the estimators using daily returns, and Panel C shows the descriptives for the estimator using monthly returns. The column denoted by "MRP (M)" shows the monthly estimator, the column denoted by "MRP (Y)" shows the annualized risk premium, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $\lambda > 0$. The rows indicate the length of the rolling estimation window used in the estimation of betas in the first stage. In Panel C one month corresponds to 20 days, thus for example the row "1200 Days" uses 60 monthly return observations.

	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val	MRP (M)	MRP (Y)	t-stat	p-val
	A: 2 Obs.				**B: 3 Obs.**				**C: 5 Obs.**			
20 Days	0.539	6.662	1.596	0.056	0.572	7.087	1.580	0.058	0.527	6.515	1.386	0.084
60 Days	0.685	8.533	1.569	0.059	0.685	8.543	1.494	0.068	0.633	7.863	1.335	0.092
120 Days	0.677	8.435	1.396	0.082	0.682	8.503	1.341	0.091	0.687	8.566	1.316	0.095
240 Days	0.316	3.856	0.615	0.270	0.316	3.857	0.595	0.276	0.270	3.283	0.503	0.308
1200 Days	-0.080	-0.956	-0.154	0.561	-0.082	-0.981	-0.152	0.560	-0.095	-1.135	-0.178	0.570
	D: 6 Obs.				**E: 13 Obs.**				**F: 26 Obs.**			
20 Days	0.603	7.483	1.529	0.064	0.660	8.213	1.496	0.068	0.652	8.112	1.427	0.078
60 Days	0.667	8.298	1.389	0.083	0.702	8.751	1.378	0.085	0.698	8.707	1.347	0.090
120 Days	0.678	8.450	1.291	0.099	0.653	8.129	1.208	0.114	0.627	7.791	1.159	0.124
240 Days	0.285	3.476	0.528	0.299	0.248	3.014	0.453	0.326	0.207	2.516	0.380	0.352
1200 Days	-0.120	-1.429	-0.222	0.588	-0.184	-2.181	-0.336	0.631	-0.217	-2.569	-0.394	0.653
	G: 39 Obs.				**H: 78 Obs.**				**I: 390 Obs.**			
20 Days	0.641	7.968	1.396	0.082	0.526	6.492	1.155	0.125	0.081	0.980	0.205	0.419
60 Days	0.660	8.211	1.281	0.101	0.509	6.279	1.023	0.154	-0.017	-0.204	-0.040	0.516
120 Days	0.573	7.095	1.076	0.142	0.408	5.003	0.807	0.210	-0.114	-1.358	-0.270	0.606
240 Days	0.169	2.051	0.315	0.376	0.025	0.302	0.049	0.480	-0.349	-4.109	-0.804	0.789
1200 Days	-0.251	-2.969	-0.457	0.676	-0.306	-3.615	-0.577	0.717	-0.439	-5.148	-0.968	0.833

4.4 Are there still 'Betting-Against-Beta' Returns?

Whether or not there is a positive market risk premium has consequences for investment strategies such as the betting-against-beta strategy. If the market risk premium is positive, i.e. if there is a positive trade-off between market risk (beta) and average return, an investor can only realize above average returns by taking above average market risk. However, if there is negative market risk premium the opposite would be true. Investors who invest in high-beta stocks would realize below average returns and investors who invest in low-beta stocks realize above average returns. In such an environment a trading strategy that takes a long position in low-beta stocks and a short position high-beta stocks would produce a positive return. Such trading strategies have been coined "betting-against-beta" strategies largely due to Frazzini and Pedersen (2014). If, however, the market risk premium is positive, betting-against-beta will yield a negative return. Therefore, we follow the same logic to compute betting-against-beta returns as we do for the estimation of market risk premiums. We show that forming betting-against-beta returns using the same beta estimates that led to positive point estimates of the market risk premium indeed lead to negative betting-against-beta returns.

We construct betting-against-beta returns by sorting stocks into high-beta (above median) and low-beta (below median) portfolios and take a long positions in the low-beta portfolio and short positions in the high-beta portfolio

$$r_t^{BAB} = \sum_{i=1}^{N} w_{i,t}^{L} r_{i,t} - \sum_{i=1}^{N} w_{i,t}^{H} r_{i,t} \tag{4.27}$$

Where $w_{i,t}^{L}$ and $w_{i,t}^{H}$ are portfolio weights for low-beta (L) and high-beta (H) portfolios, such that $\sum_{i=1}^{N} w_{i,t}^{L} = \sum_{i=1}^{N} w_{h,t}^{H} = 1$. Applying the same weighting scheme as Frazzini and Pedersen (2014), a stocks receives a larger weight the further its beta is from the median beta. Therefore, we compute portfolio weights as

$$w_{i,t}^{L} = \begin{cases} \eta_{i,t-1}(\overline{rank}_{t-1} - rank_{i,t-1}) & \text{if } rank_{i,t-1} < \overline{rank}_{t-1} \\ 0 & otherwise \end{cases}$$
$$w_{i,t}^{H} = \begin{cases} \eta_{i,t-1}(rank_{i,t-1} - \overline{rank}_{t-1}) & \text{if } rank_{i,t-1} > \overline{rank}_{t-1} \\ 0 & otherwise \end{cases} \tag{4.28}$$

where $rank_{i,t-1} = rank(\beta_{i,t-1})$ denotes the rank of $\beta_{i,t-1}$ in the sorted cross-section. $\overline{rank}_{t-1} = (1/N)\sum_{i=1}^{N} rank_{i,t-1}$ denotes the average rank. Thus, $w_{i,t}^{L} > 0$ for stocks with a beta below the median beta and zero for all other stocks. Accordingly, $w_{i,t}^{H} > 0$ for stocks with a beta above the median beta and zero for all other stocks. In either case the weight increases the further the stock's beta is from the median. To ensure weights each add up to one we multiply the differences between individual rank and average rank by a normalization factor $\eta_{i,t-1} = 2/ \mid rank_{i,t-1} - \overline{rank}_{t-1} \mid$. To mitigate the influence of outliers in estimated betas we shrink each beta towards the theoretical cross-sectional mean $\bar{\bar{\beta}} = 1$, i.e. we have $\tilde{\hat{\beta}}_i = 0.6 * \hat{\beta}_i + 0.4 * \bar{\bar{\beta}}$ (see Frazzini and Pedersen (2014), Elton, Gruber, Brown, and Goetzmann (2003), Vasicek (1973)). For the trading strategy in equation (4.27) we expect $r_T^{BAB} > 0$ if $\lambda_T \leq 0$ and $r_T^{BAB} \leq 0$ if $\lambda_T > 0$. For each estimation of the market risk premium in sub-chapter 4.3.2 to sub-chapter 4.3.4 we compute r_T^{BAB} using equation (4.27) and report the average returns of the trading strategy using the arithmetic mean.

We show the average BAB returns for the sample period from August 10, 1926 to December 31, 2013 in Table 4.4.1. The Table is constructed in the same way as Table 4.3.5 to facilitate an easier comparison. We find positive BAB returns in Table 4.4.1 whenever the estimated market risk premium in Table 4.3.5 is negative and vise versa. For the total CRSP cross-section and betas estimated on rolling windows of daily returns (Panel A) we find an annualized average BAB return of approximately 1.585%. This is consistent with BAB returns for the U.S. stock market reported in Frazzini and Pedersen (2014).[13]

[13]While we use the same weighting scheme and definitions for high- and low-beta portfolios as Frazzini and Pedersen (2014), they construct the trading strategy returns slightly differently to ensure that their BAB strategy is market neutral. Without loss in generality we deviate form a market neutral BAB strategy. If the market risk premium is positive a market neutral BAB strategy as in Frazzini and Pedersen (2014) would

Recall that rows denoted by "CRSP Total" shows the statistics for the total CRSP sample. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly. As we move from the total CRSP sample to the highest liquidity sample "Min. Four Top" the average BAB returns become more negative. At the same time the corresponding market risk premiums in Table 4.3.5 become more positive.

Table 4.4.1: Average BAB returns August 1926 to December 2013

The table shows the average BAB returns for the sample period August 1926 to December 2013. The BAB returns are constructed using equation (4.27). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A uses daily returns of a time period of 240 days and Panel B uses monthly returns of a time period of 60 months. In the second step these betas used to construct the betting-against-beta (BAB) returns, with the immediately following 20 days as holding period. The column denoted by "BAB (M)" shows the monthly average return, the column denoted by "BAB (Y)" shows the annualized average return, the column "t-stat" shows the t statistic for the monthly average return, and the column "p-val" is computed using the alternative hypothesis $r^{BAB} > 0$. *The rows denoted by "CRSP" shows the statistics for the total CRSP sample. The rows denoted by "No Top" denote the statistics for the sample that does only contain stocks that are never in the top tercile of liquidity measures. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly.*

	A: 240 Days				**B: 60 Months**			
	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val
CRSP Total	0.131	1.585	0.941	0.173	-0.231	-2.735	-1.512	0.935
Min. One Top	0.003	0.037	0.021	0.492	-0.364	-4.277	-2.376	0.991
Min. Two Top	-0.134	-1.600	-0.883	0.811	-0.387	-4.544	-2.558	0.995
Min. Three Top	-0.206	-2.440	-1.372	0.915	-0.439	-5.143	-2.956	0.998
Min. Four Top	-0.366	-4.306	-2.327	0.990	-0.558	-6.498	-3.702	1.000

We obtain the same patter for Table 4.4.2. Whenever the market risk premium in Table 4.3.6 is positive the average BAB returns in Table 4.4.2 are negative and vise versa. In the most liquidly traded samples – the 'Min. Four Top sample and the S&P 500 sample – the BAB returns are negative and rather large in absolute magnitude. These results are consistent with findings reported in Novy-Marx and Velikov (2016) and Li, Sullivan, and

earn the risk-free rate and positive return above the risk-free rate if the market risk premium is negative.

Garcia-Feijoo (2014) who show that betting-against-beta returns are concentrated in small and illiquid stocks. In samples of liquid and large cap stocks we obtain positive market risk premiums and betting-against-beta yields a negative average return. This suggests that the returns produced by betting-against-beta strategies can be explained by an illiquid premium not by behavioral biases. In Tables 4.4.3 to 4.4.5 we report average BAB returns in the same

Table 4.4.2: Average BAB returns January 1999 to December 2013

The table shows the average BAB returns for the sample period January 1999 to December 2013. The BAB returns are constructed using equation (4.27). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A uses daily returns of a time period of 240 days and Panel B uses monthly returns of a time period of 60 months. In the second step these betas used to construct the betting-against-beta (BAB) returns, with the immediately following 20 days as holding period. The column denoted by "BAB (M)" shows the monthly average return, the column denoted by "BAB (Y)" shows the annualized average return, the column "t-stat" shows the t statistic for the monthly average return, and the column "p-val" is computed using the alternative hypothesis $r^{BAB} > 0$. The rows denoted by "CRSP" shows the statistics for the total CRSP sample. The rows denoted by "No Top" denote the statistics for the sample that does only contain stocks that are never in the top tercile of liquidity measures. The rows denoted by "Min. One Top" denotes the statistics containing stocks that are in the top tercile of at least one liquidity measure. The rows "Min. Two Top", "Min. Three Top", and "Min. Four Top" used samples constructed accordingly. The row denoted "S&P 500" denotes the S&P500 index universe sample.

	A: 240 Days				**B: 60 Months**			
	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val
CRSP Total	0.360	4.411	0.852	0.198	-0.500	-5.835	-1.125	0.869
Min. One Top	0.439	5.395	0.929	0.177	-0.555	-6.459	-1.094	0.862
Min. Two Top	0.173	2.093	0.337	0.368	-0.710	-8.197	-1.460	0.927
Min. Three Top	-0.168	-1.993	-0.331	0.630	-0.743	-8.557	-1.587	0.943
Min. Four Top	-0.368	-4.327	-0.708	0.760	-0.761	-8.756	-1.625	0.947
S&P 500	-0.432	-5.060	-0.903	0.816	-0.583	-6.779	-1.305	0.903

way we report market risk premiums in Tables 4.3.8 to 4.3.11. Consistently over all tables the average BAB returns are negative if estimated market risk premiums are positive and vise versa. This suggests that positive BAB returns are driven by the same statistical biases that also drive negative market risk premiums.

Table 4.4.3: Average BAB returns and stock return frequency

The table shows average BAB returns for the sample period January 1999 to December 2013. BAB returns are calculated using equation (4.27). In a first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A rolling estimation windows of 12 months and Panel B uses 60 months. In a second step these betas are used to obtain BAB returns using equation (4.27). The holding period are the immediately following 20 days. The column denoted by "BAB (M)" shows the monthly average returns, the column denoted by "BAB (Y)" shows the annualized average BAB returns, the column "t-stat" shows the t statistic for the monthly averages, and the column "p-val" is computed using the alternative hypothesis that the average BAB return is $>$ 0. The rows denoted by "Intra-Day" shows the statistics if the beta estimator is based on intra-day returns within the estimation window, here using 2 return observations per day. The rows denoted by "Daily", "Weekly", "Bi-Weekly", "Monthly" denote the statistics if the beta estimator is based in daily, weekly, bi-weekly, or monthly returns respectively within the estimation window.

	A: 12 Months				B: 60 Months			
	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val
Intra-Day	-0.149	-1.776	-0.270	0.606	-0.045	-0.541	-0.097	0.538
Daily	-0.432	-5.060	-0.903	0.816	-0.284	-3.358	-0.639	0.738
Weekly	-0.408	-4.789	-0.875	0.809	-0.433	-5.070	-0.930	0.823
Bi-Weekly	-0.309	-3.642	-0.726	0.766	-0.522	-6.084	-1.135	0.871
Monthly	-0.409	-4.805	-1.026	0.847	-0.583	-6.779	-1.305	0.903

Table 4.4.4: Average BAB returns and beta estimation window length

The table shows shows the estimated risk premium for the sample period January 1999 to December 2013. BAB returns are calculated using equation (4.27). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A shows the descriptives for the estimators using high-frequency returns. Panel B shows the descriptives for the estimators using daily returns, and Panel C shows the descriptives for the estimator using monthly returns. The column denoted by "BAB (M)" shows the monthly estimator, the column denoted by "BAB (Y)" shows the annualized average BAB returns, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $r^{BAB} > 0$. *The rows indicate the length of the rolling estimation window used in the estimation of betas in the first stage. In Panel C one month corresponds to 20 days, thus for example the row "1200 Days" uses 60 monthly return observations.*

	A: High Freq.				**B: Daily**				**C: Monthly**			
	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val
20 Days	-0.616	-7.145	-1.248	0.893	-0.608	-7.059	-1.487	0.931	–	–	–	–
60 Days	-0.658	-7.614	-1.207	0.886	-0.505	-5.890	-1.097	0.863	-0.317	-3.741	-1.246	0.893
120 Days	-0.519	-6.054	-0.950	0.828	-0.495	-5.780	-1.057	0.854	-0.487	-5.695	-1.448	0.925
240 Days	-0.149	-1.776	-0.270	0.606	-0.432	-5.060	-0.903	0.816	-0.409	-4.805	-1.026	0.847
1200 Days	-0.045	-0.541	-0.097	0.538	-0.284	-3.358	-0.639	0.738	-0.583	-6.779	-1.305	0.903

Table 4.4.5: Average BAB returns from Intra-Day Stock Returns

The table shows average BAB returns for the sample period January 1999 to December 2013. BAB returns are calculated using equation (4.27). In the first step betas are estimated by a time series regression of the value weighted market return computed by CRSP on returns of single stocks. Panel A shows the descriptives for the estimators using high-frequency returns. Panel B shows the descriptives for the estimators using daily returns, and Panel C shows the descriptives for the estimator using monthly returns. The column denoted by "BAB (M)" shows the monthly estimator, the column denoted by "BAB (Y)" shows the annualized average BAB returns, the column "t-stat" shows the t statistic for the monthly estimator, and the column "p-val" is computed using the alternative hypothesis $r^{BAB} > 0$. The rows indicate the length of the rolling estimation window used in the estimation of betas in the first stage. In Panel C one month corresponds to 20 days, thus for example the row "1200 Days" uses 60 monthly return observations.

	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val	BAB (M)	BAB (Y)	t-stat	p-val
	A: 2 Obs.				**B: 3 Obs.**				**C: 5 Obs.**			
20 Days	-0.616	-7.145	-1.248	0.893	-0.614	-7.129	-1.225	0.889	-0.639	-7.400	-1.247	0.893
60 Days	-0.658	-7.614	-1.207	0.886	-0.667	-7.722	-1.213	0.887	-0.603	-6.999	-1.091	0.862
120 Days	-0.519	-6.054	-0.950	0.828	-0.534	-6.223	-0.969	0.833	-0.482	-5.631	-0.871	0.807
240 Days	-0.149	-1.776	-0.270	0.606	-0.151	-1.793	-0.273	0.607	-0.092	-1.101	-0.168	0.567
1200 Days	-0.045	-0.541	-0.097	0.538	-0.024	-0.285	-0.052	0.521	0.008	0.098	0.018	0.493
	A: 6 Obs.				**B: 13 Obs.**				**C: 26 Obs.**			
20 Days	-0.665	-7.691	-1.284	0.900	-0.612	-7.107	-1.167	0.878	-0.551	-6.417	-1.068	0.857
60 Days	-0.617	-7.163	-1.113	0.866	-0.567	-6.597	-1.023	0.846	-0.490	-5.724	-0.907	0.817
120 Days	-0.484	-5.656	-0.876	0.809	-0.431	-5.052	-0.788	0.784	-0.328	-3.862	-0.618	0.731
240 Days	-0.092	-1.099	-0.168	0.567	-0.036	-0.434	-0.067	0.527	0.064	0.773	0.123	0.451
1200 Days	0.026	0.312	0.057	0.477	0.082	0.984	0.185	0.427	0.129	1.555	0.302	0.382
	A: 39 Obs.				**B: 78 Obs.**				**C: 390 Obs.**			
20 Days	-0.515	-6.013	-1.015	0.844	-0.363	-4.274	-0.753	0.774	0.123	1.486	0.311	0.378
60 Days	-0.424	-4.967	-0.803	0.788	-0.246	-2.907	-0.495	0.690	0.236	2.869	0.575	0.283
120 Days	-0.262	-3.095	-0.506	0.693	-0.084	-1.003	-0.174	0.569	0.368	4.501	0.919	0.180
240 Days	0.114	1.377	0.224	0.412	0.239	2.904	0.503	0.308	0.493	6.080	1.241	0.108
1200 Days	0.154	1.867	0.368	0.357	0.219	2.665	0.556	0.290	0.379	4.650	1.140	0.128

4.5 Conclusion

In summary, we show analytically that price measurement error, sampling error, and time-series variation in betas can lead to heavily attenuated estimates of the market risk premiums and to biased statistical inference in the standard OLS-based two-pass Fama-MacBeth estimation. The non-trivial interrelation between price measurement error, sampling error, and time-series variation in betas complicates the well-known errors-in-variables problem in the second stage of the Fama-MacBeth procedure and can lead not only to estimated market risk premiums close to zero but can even produce negative estimates if the true market risk premium is positive. Empirically, we show that negative market risk premiums predominantly occur in samples with rather low average liquidity and especially in more recent sample periods if long estimation windows are chosen. In our long-run sample from 1926 to 2013 price measurement error appears to be the dominant source of bias, while in the more recent sample period from 1999 to 2013, time-series variation in betas seems to increase in relative importance. This could be the result of increased, faster trading. Our results also indicate that intra-day returns are only of limited help in the estimation of betas and market risk premiums as too high sampling frequencies of returns increase noise in estimated betas.

Overall, we find that the estimation of market risk premiums using the standard OLS based two-pass Fama-MacBeth procedure is rather sensitive to choices that were made in the estimation of betas regarding the estimation window length and the sampling frequency of returns within the estimation window. Furthermore, estimates of the market risk premium are heavily affected by the average liquidity and price staleness in the employed sample. These sensitivities are often neglected in empirical studies. The sensitivity of OLS estimates to price measurement error, sampling error, and time-series variation in betas suggest that researchers should apply OLS estimation of betas carefully, as it is difficult to estimate betas precisely in this way. More sophisticated estimators such as the option-implied beta estimator in Buss and Vilkov (2012) or the valuation based shrinkage estimator in Cosemans et al. (2016), are likely to be preferable in most datasets.

Our empirical and analytical results suggest that these more sophisticated estimation techniques do not simply correct for measurement and sampling errors but help in incorporating additional information that improves accuracy and precession of beta estimates in situations that produce noisy OLS estimates, and therefore alleviate attenuation bias in the estimation of market risk premiums. For example, Hu (2014) provides evidence that option prices convey significant information about their underlying stocks. Betas constructed from option-implied variances and covariances incorporate this additional information. The approached employed by Cosemans et al. (2016) combines stock return information with valuation information to estimate individual firm betas. For example, Cochrane (1991) suggests a production-based asset pricing model that extends the traditional consumption-based view by tying stock returns to investment returns. Zhang (2005) shows that due to costly reversibility of capital investments, value firms have countercyclical betas and growth firms have procyclical betas. The estimation technique proposed by Cosemans et al. (2016) efficiently combines consumption-based asset pricing—i.e. beta information from stock returns—with production-based asset pricing—i.e. beta information from firm fundamentals—by finding the weighted average between the two, attributing to each individual stock the higher weight on the more informative beta. Our results indicate that the estimator would attribute higher weight to more liquid stocks as they are estimated more precisely.

We also show that positive betting-against-beta returns are driven by price measurement error, sampling error, and time-series variation in betas. The concentration of betting-against-beta returns in rather illiquid sub-samples suggests that the good past performance of betting-against-beta strategies is more likely driven by an illiquidity premium than by exploiting mispricing. We also find that average betting-against-beta returns exhibit the same sensitivity towards the length of the estimation window and the sampling frequency of returns in the estimation of betas as market risk premiums. Without clear criteria regarding these choices and given the concentration of positive returns of betting-against-beta strategies in illiquid stocks investors should approach betting-against-beta with caution.

5

Research Impact

The research presented in this thesis addresses three questions that are of interest to both academic and public audiences.

Chapter 2, titled *"Credit Supply: Are there negative spillovers from banks' proprietary trading?"*, investigated the impact of banks' proprietary trading on credit supply. Due to their systemic importance and due to the implicit and explicit government guarantees that banks enjoy, banks' business practices are the frequent topic of discussion around economic policy and are subject to significant public scrutiny. In particular, the proprietary trading activities of banks have come under great scrutiny since the financial crisis that began in 2007. The Volcker Rule in the US, the Vickers Report in the UK, and the Liikanen and European Commission proposals in the EU all aim at a limitation of risks believed to emanate from banks' trading activities by strictly separating trading from commercial banking business. The concern reflected in these rules is that banks take on large risky bets, while relying on implicit or explicit government guarantees for cheap funding and threatening to reduce their credit supply to the real economy. Indeed, the results in Chapter 2 show that compared to non-trading banks, trading banks reduce credit supply to corporate borrowers by 19% and reduce it even further during periods of crisis. Moreover, the findings in Chapter 2 also indicate that these spillovers from trading to credit supply have adverse consequences for the real economy as firms that borrow from trading banks have a lower ability to invest in capital and expand their workforce. However, these real economic effects are rather weak compared to the decline in credit supply. Thus, when discussing new regulations regarding banks' proprietary trading, policymakers and regulators should consider the possibility that borrowers may have the ability to compensate for reduced bank credit supply through other means of financing. Additionally, Chapter 2 contributes to the policy discourse by showing that even for globally active banks the relation between proprietary trading and credit supply is not homogeneous across country borders. More stringent regulations on banks' proprietary trading in the US compared to other economies' regulations appear justified, as US banks reduce credit supply more heavily if they engage in proprietary trading. Overall, the results in Chapter 2 highlight that policymakers and regulators should initiate careful cost-benefit analyses of the implementation of bans on proprietary trading or separation of proprietary trading from commercial banking. Chapter 2 is a first step towards such cost-benefit analyses.

Chapter 3 addressed another aspect of banks' business practices that is controversially discussed whereas the chapter's central question is posed in its title: *"Is Reported Derivative Use Informative About Risk-Taking?"* This question directly derives from a more fundamental question that is relevant for everyone who analyzes accounting data, no matter whether it is for academic research, supervisory purpose or investment choices: Do reported accounting figures provide accurate information about a firm's economic reality, given that managers can often exercise considerable discretion when compiling balance sheets? Chapter 3 demonstrates a possible approach to combining corporate finance theory with statistical data analysis to obtain an estimate of a bank's economic reality of derivatives use that can be compared with the reported reality. Supervisory authorities and auditors could use similar approaches for plausibility checks of accounting figures in cases in which managers have significant discretion when compiling the figures or in cases of voluntary reporting.

Since investors and regulators need to rely on what is observable about banks' economic reality, it is important to design reporting rules such that banks have incentives to accurately report how they use derivatives. The results in Chapter 3 indicate that in many situations, banks under-report the extent of their hedging activities. While this may be seen as desirable from a regulatory perspective, as it leads to an overestimation rather than an underestimation of potential risks in the banking sector, it also carries the risk of choking financial innovation or deters banks from effective hedging opportunities. Therefore, the chapter's results suggest that current hedge accounting rules should be simplified.

Finally, Chapter 4, titled *"Nobody Knew that Measurement Error Could be so Complicated: A Note on Estimating Betas & Market Risk Premiums"*, showed that a negative or flat market risk premium (i.e. a low-beta anomaly) can be explained by statistical biases rather than behavioral bias or market frictions. Therefore, the low-beta anomaly appears to be an artifact of mis-measuring market risk. In the introductory chapter of this thesis (Chapter 1) the analysis presented in Chapter 4 is motivated from a banking perspective. Since banks largely depend on financial markets to raise capital to meet regulatory capital requirements, the presence of a low-beta anomaly would directly impact banks' cost of capital. Put in simplified terms the low-risk anomaly constitutes that investors require compensation for excepting less rather than more risk. In such an environment, banks would face higher cost of capital as they increase equity and become saver as a consequence. Therefore, the results in Chapter 4 have direct consequences for policy discussions around capital requirements for banks. The presence of a low-beta anomaly should not be accepted as a potential excuse against higher equity capital.

Besides this more indirect implication, the results in Chapter 4 also have direct implications for the investors applying betting-against-beta strategies or investing in low-beta ETFs. The good performance of such strategies and ETFs is not due to an out-performance of low-beta stocks compared to high-beta stocks but is due to an illiquidity premium and mis-measured betas of illiquid stocks. Betas are difficult to estimate precisely. Investors using betas as an ingredient in their investment strategies should pay careful attention to the fundamental statistical properties of the beta estimators, even if the estimator is as simple as OLS, to avoid misleading conclusions.

6
Concluding Remarks

THIS THESIS summarizes my research since October 2015, conducted as a PhD candidate in the Finance Department at Maastricht University. The approaches used and the specific questions that are asked in each chapter differ, and each of the chapters may be seen as contributing to another stream of literature. Nevertheless, each chapter can be motived from a banking perspective and has some implications for investors and regulators in the banking sector.

The rather broad scope of this thesis came to be largely due to my own curiosity and wide interests. I admittedly never made a plan as to what topics to include in my thesis, rather I followed my curiosity. From the perspective of my personal motivation, most of the time this was driven by an interest in the application of certain econometric methods or the wish to work with specific datasets. I started working on Chapter 2 because I wanted to work with microdata containing bank-borrower relations; Chapter 3 was largely motivated by my wish to work on an economic application of the expectation-maximization algorithm after my supervisor, Jaap, showed me some of his own work on the topic. And Chapter 4 emerged out of an interest in and curiosity about the beta estimators suggested in Buss and Vilkov (2012) and Cosemans et al. (2016) after another of my supervisors, Paulo, introduced me to this line of research. Approaching a PhD thesis in such a way has its advantages and disadvantages. For each of the chapters, I had to work with rather different datasets and different methods and had to dive into different streams of literature. The advantage of this is that it taught me a lot about how to merge, create, and prepare datasets from various sources. It definitely broadened my horizons and taught me that the borders we draw between many subjects or streams of literature are in fact more blurry than initially perceived. Of course, the great disadvantage of writing a thesis with a rather broad scope is that I never reached the level of specialization in any of the particular topics that I would have done if I were to only focus on one of them.

Chapter 2 investigated the question: "Do banks that heavily engage in proprietary trading reduce credit supply in times of crisis more than their peers less heavily engaged in proprietary trading?" In our analysis we answered this question using a global dataset containing information on loans granted by 136 leading banks to a wide range of corporate borrowers between 2003 and 2016. We found that banks with greater trading expertise supply less credit during stable times and even less during crisis times. Compared to non-trading banks, trading banks reduce credit supply by 19% plus an additional 3.25% per unit increase in the Financial Stress Indicator of the US Office of Financial Responsibility. Both effects are consistent with theoretical predictions (see Shleifer and Vishny (2010), Diamond and Rajan (2011), Boot and Ratnovski (2016)) and are in line with previous empirical evidence derived from a one-country sample (see Abbassi et al. (2016)). Additionally, we demonstrated that banks engaged in trading also charge higher prices for their loans. Moreover, we showed that the global dimension of our analysis matters. The double effect of trading banks reducing credit supply during crisis and non-crisis periods can be attributed to US banks. International banks are special as they only reduce their credit supply during crises. From a theoretical point of view this suggests that between US banks and international banks there are two different channels at work, both leading to lower credit supply. The theoretical model suggested in Boot and Ratnovski (2016) predicts that banks with trading expertise allocate scarce funds to scalable short-term securities trading rather than non-scalable long-term relationship lending activities, thus leading to lower credit supply. This channel appears to be at work in US banks but not in international banks. On the other hand, Shleifer and Vishny (2010) and Diamond and Rajan (2011) argue that banks with trading expertise redirect funds away from lending to trading during periods of crisis as the returns from investing in distressed assets are higher than returns from lending. This channels appears to be at work both in US and international banks. These differences help in the assessment of differences in the regulatory frameworks regarding proprietary trading in the US and, e.g., Europe, with US regulations being significantly more restrictive. Further exploiting our global sample, we also found that while trading banks provide less credit overall they tend to provide slightly more credit abroad. However, during a crisis, trading banks also cut their foreign lending to a greater extent than their non-trading peers. Finally, we showed that these spillovers from trading to credit supply have adverse consequences for the real economy as firms have reduced ability to invest in capital and expand their workforce. In particular, this last point adds important information to the debate around the new regulations of banks' proprietary trading as it shows that there are externalities of proprietary beyond excessive risk-taking by banks. Therefore, this constitutes a first step towards a cost-benefit analysis of regulations that restrict banks in their proprietary trading operations. However, our analysis also showed that real economic impact, while present, is limited. Since our sample consists of stock exchange-listed borrowers, this suggests that these borrowers have the ability to compensate the reduced bank credit supply through other sources of funding. An extension of our analysis that also includes non-listed borrowers would likely be a fruitful avenue for future research. However, data on non-listed firms is difficult to obtain and is often only available for a limited number of European economies. Overall, our results suggest that the recent regulatory initiatives to separate trading from commercial banking activities such as lending are generally well advised, as banks that engage heavily in proprietary trading reduce their credit supply relative to other banks. Moreover, we showed that a global perspective matters for the assessment of spillovers from trading to lending.

Chapter 2 highlighted the importance of reliable and detailed data to assess the impact of banking regulation. Besides accounting data on banks and borrowers Chapter 2 required two other ingredients that are difficult to come by: a measure of banks' proprietary trading activities and microdata on banks' lending activities. Any cost-benefit analysis of restrictions to banks trading businesses would need such data as well. Proprietary trading activities of banks are hard to measure. However, at the European level the European Central Bank (ECB) already collects detailed data on securities holdings of European banks in its Securities Holdings Statistics (SHS) database. This allows for a more detailed analysis of banks' trading activities as this database records prices, quantities, time of purchase and sale of securities at the security-level per bank. The possibilities to use this data to answer the questions raised in Chapter 2 are, however, limited even if one would focus only on Europe. This is due to the low availability of microdata on banks' lending activities. Credit register data is particularly useful in this context as credit registers allow the tracking of the development of loans over time, something that is not possible with the LPC DealScan data used in Chapter 2. However, credit registers do not exist in many countries, and for the countries in which they do exist, reporting requirements may differ substantially. Furthermore, national credit registers limit the scope of any analysis to a single country. Since September 2018, the ECB is collecting data from European banks for its AnaCredit database, a database that is comparable to a multi-country credit register.[1] Combining the SHS and AnaCredit databases with European-level databases on listed and non-listed banks and borrowers would allow the overcoming of many of the limitations shown in Chapter 2. Another promising avenue for research on the issues raised in Chapter 2 would be an analysis of loan applications rather than granted loan volumes. This would allow us to model the probability of a firm being able to obtain a loan from trading and non-trading banks.

Chapter 3 investigated trading as compared to hedging in the context of derivatives use by banks. Specifically, the chapter attempted to answer the question: "Is Reported Derivative Use Informative About Risk-Taking?" Bank managers can exercise considerable judgment when applying hedge accounting rules to designate derivatives as trading or hedging instruments. If a bank decides to designate a derivative as a hedging instrument it has to provide extensive evidence regarding the effectiveness of the hedge relation and has to comply with extensive documentation requirements. In surveys, banks claim that this leads to under-reporting of the extent of hedging activities as the bureaucratic burden of designating a derivative as hedge often outweighs the benefit. Together with the overall

[1] AnaCredit is an abbreviation for *analytical credit datasets*. See `www.ecb.europa.eu/stats/money_credit_banking/anacredit` for more information.

extraordinarily high complexity of derivatives use, reporting this creates a situation in which it is unclear whether reported derivatives use helps regulators and investors in the assessment of banks' derivatives use and whether this derivatives use is related to banks' risk attitudes. Employing a latent class regression model, we showed in Chapter 3 that reported derivatives use under current hedge accounting rules is only weakly related to bank risk. Therefore, reported derivatives use does not provide investors with an indication of whether a bank applies a derivatives strategy focused on either hedging various risks or focused in speculative trading. While reported derivatives use tends to underestimate the extent of hedging activities in the banking sector, it also appears to underestimate the risk of those banks reporting most their derivatives as trading instruments. Creating reporting rules for derivatives, such that there is an incentive for banks to under-report hedging, may be reasonable from a regulatory perspective as it results in 'worst-case' picture of banks' derivatives use. However, the Accounting Standards Board's (ASB) Statement of Principles published in 1999 declares that "*the objective of financial statements is to provide information [...] useful for assessing stewardship [...] and for making economic decisions*". Our results suggest that current reporting rules for derivatives use are inconsistent with this principle.

Thus, Chapter 3 is an attempt to model the unobserved *true* derivatives use of banks through a latent class regression model. Following the approach in Bos et al. (2010a) and Bos et al. (2010b) we allow banks to move across classifications as hedgers and traders by treating bank-quarter observations in the analysis as independent. A potentially insightful avenue for future research would be to explicitly model changes in latent class membership over time, for example through a Hidden Markov Model. Furthermore, the analysis in Chapter 3 takes a bank-level view rather than a securities-level view. Therefore, the classifications indicate a derivatives use strategy of a bank but does not allow any conclusions about any particular derivate. Generally, the latent class regression model applied in Chapter 3 could also be adapted to a securities-level analysis that would allow for a classification of every individual derivative instrument into hedging and trading. Such an analysis would be particularly useful for supervisory authorities or auditors. However, it would require extensive microdata on cash flow generated by the individual derivatives and the cash flows originating from e.g. loan portfolios, other investments, or from the banks' own debt obligations.

Finally, Chapter 4 took a step back from banking and investigated how risk is priced in the stock market, or more precisely, how the way market risk and its price are measured affect the conclusions about market efficiency. The central point of the chapter is that market risk defined as a stocks beta is difficult to estimate precisely. We argue that the most common estimation techniques for beta are flawed, as the resulting estimates are impacted by three sources of error: price-measurement error, sampling error and time-series variation. As a result, the typical beta estimate is a poor predictor of the realized beta, causing spurious asset pricing 'anomalies'. We showed that the low-beta anomaly is a result of this statistical bias, rather than behavioral bias or market frictions. We demonstrated that by adapting the estimation technique to these three sources of error, we obtained a significant positive market risk premium between 6% and 8% per annum. The solution is simple: minimize sampling error through more frequent sampling; minimize the error due to time-series variation by estimating betas over a small window; and minimize price error measurement by maintaining a sample of liquid stocks. We also show that positive betting-against-beta returns are driven by price-measurement error, sampling error, and time-series variation in betas. Concentration of betting-against-beta returns in rather illiquid sub-samples suggests that the good past performance of betting-against-beta strategies is more likely driven by an illiquidity premium than by exploiting mispricing. We also find that average betting-against-beta returns exhibit the same sensitivity towards the length of the estimation window and the sampling frequency of returns in the estimation of betas as market risk premiums. Without a clear criterion regarding these choices and given the concentration of positive returns of betting-against-beta strategies in illiquid stocks, investors should approach betting-against-beta with caution.

References

Abbassi, P., Iyer, R., Peydro, J.-L., and Tous, F. R. Securities trading by banks and credit supply: Micro-evidence from the crisis. *Journal of Financial Economics*, 121:569–594, 2016.

Acharya, V. V., Brownlees, C., Engle, R., Farazmand, F., and Richardson, M. Measuring Systemic Risk. In Roggi, O. and Altman, E. I., editors, *Managing and Measuring Risk*, volume 5 of *World Scientific Series in Finance*, pages 99–122. World Scientific Publishing, 2013.

Acharya, V. V., Eisert, T., Eufinger, C., and Hirsch, C. Real Effects of the Sovereign Debt Crisis in Europe: Evidence from Syndicated Loans. *Review of Financial Studies*, 31(8): 2855–2896, 2018.

Admati, A. R., DeMarzo, P. M., Hellwig, M. F., and Pfleiderer, P. Fallacies, Irrelvant Facts, and Myths in the Discussion of Capital Regulation: Why Bank Equity is Not Expensive. *Working Paper*, 2011.

Ahmed, A. S., Kilic, E., and Lobo, G. J. Effects of SFAS 133 on the Risk Relevance of Accounting Measures of Banks' Derivative Exposures. *The Accounting Review*, 86(3): 769–804, 2011.

Allen, F. and Carletti, E. The Role of Banks in Financial Systems. In Berger, A. N., Wilson, J. O. S., and Molyneux, P., editors, *The Oxford Handbook of Banking*, chapter 2, pages 37–57. Oxford University Press, Oxford, 2009.

Allen, F. and Gale, D. Financial markets, intermediaries, and intertemporal smoothing. *Journal of Political Economy*, 105(3):523–546, 1997.

Allen, F., Chui, M. K., and Addaloni, A. Financial Systems in Europe, the USA, and Asia. *Oxford Review of Economic Policy*, 20(4):490–508, 2004.

Altman, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 23(4):589–609, 1968.

Altman, E. I. Predicting financial distress of companies: revisiting the Z-score and Zeta models. *NYU Working Paper*, 2000.

Altman, E. I. Corporate Distress Prediction Models in a Turbulent Economic and Basel II Environment. *NYU Working Paper No. S-CDM-02-11*, 2002.

Amihud, Y. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5:31–56, 2002.

Ang, A. *Asset Management: A Systemic Approach to Factor Investing*. Oxford University Press, New York, 2014.

Ang, A., Liu, J., and Schwarz, K. Using Stocks or Portfolios in Tests of Factor Models. *Working Paper*, 2018.

Argote, L. and Miron-Spektor, E. Organizational Learning: from experience to knowledge. *Organization Science*, 22(5):1123–1137, 2011.

Arping, S. Propreity Trading and the Real Economy. *Tinbergen Institute Discussion Paper*, 2013.

Asparouhova, E., Bessembinder, H., and Kalcheva, I. Liquidity biases in asset pricing tests. *Journal of Financial Economics*, 96:215–237, 2010.

Asparouhova, E., Bessembinder, H., and Kalcheva, I. Noisy Prices and Inference Regarding Returns. *Journal of Finance*, 68(2):665–714, 2013.

Avraham, D., Selvaggi, P., and Vickery, J. A Structural View of U.S. Bank Holding Companies. *FRBNY Economic Policy Review*, (July), 2012.

Baillie, R. T. and DeGennaro, R. P. Stock Returns and Volatility. *Journal of Financial and Quantitative Analysis*, 25(2):203–214, 1990.

Baker, M. Capital Market-Driven Corporate Finance. *Annual Review of Financial Economics*, 1:181–205, 2009.

Baker, M. and Wurgler, J. Do Strict Capital Requirenments Raise the Cost of Capital? Bank Regulation, Capital Structure, and the Low-Risk Anomaly. *American Economic Review: Papers & Proceedings*, 105(5):315–320, 2015.

Baker, M., Bradley, B., and Wurgler, J. Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly. *Financial Analysts Journal*, 67(1):40–54, 2011.

Bali, T. G., Engle, R. F., and Tang, Y. Dynamic Conditional Beta Is Alive and Well in the Cross Section of Daily Stock Returns. *Management Science*, 63(11):3760–3779, 2017.

Bandi, F., Pirino, D., and Reno, R. Excess Idle Time. *Econometrica*, 85(6):1793–1846, 2017.

Bandi, F., Pirino, D., and Reno, R. Systemic staleness. *Working Paper*, 2018a.

Bandi, F., Pirino, D., and Reno, R. Zeros. *Working Paper*, 2018b.

Bao, J., O'Hara, M., and Zhou, X. A. The Volcker Rule and corporate bond market making in times of stress. *Journal of Financial Studies*, Forthcoming.

Bekaert, G., Harvey, C. R., and Lundblad, C. Liquidity and Expected Returns: Lessons from Emerging Markets. *Review of Financial Studies*, 20(6):1783–1831, 2007.

Berry, J. M. Divided on Derivatives. *The Washington post*, 2003. URL `https://www.washingtonpost.com/archive/business/2003/03/06/divided-on-derivatives/271810fb-e3b5-48ef-a153-80bc26ccb9ff/?noredirect=on&utm_term=.3c2db325ebc8`.

Bertoletti, M., Friel, N., and Rastelli, R. Choosing the number of clusters in a finite mixture model using an exact integrated completed likelihood criterion. *METRON*, 73(2): 177–199, 2015.

Bharath, S. T. and Shumway, T. Forecasting default with the Merton Distance to Default Model. *The Review of Financial Studies*, 21(3), 2008.

Biernacki, C., Celeux, G., and Govaert, G. Assessing a mixture model for clustering with the completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725, 2000.

Biernacki, C., Celeux, G., and Govaert, G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, 41:561–575, 2003.

Black, F., Jensen, M. C., and Scholes, M. The Capital Asset Pricing Model: Some Empirical Tests. *Studies in the Theory of Capital Markets*, 1972.

Blanchard, O. J., Das, M., and Faruqee, H. The Initial Impact of the Crisis on Emerging Market Countries. *Brookings Papers on Economic Activity*, pages 263–307, 2010.

Blume, M. E. Portfolio Theory: A Step Towards Its Pratical Application. *Journal of Business*, 43(2):152–173, 1970.

Blume, M. E. Betas and their regression tendencies. *Journal of Finance*, 30(3):785–795, 1975.

Blume, M. E. and Stambaugh, R. F. Biases in Computed Returns: An Application to the Size Effect. *Journal of Financial Economics*, 12:387–404, 1983.

Blyth, C. R. On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*, 67(338):364–366, 1972.

Bolton, P., Mehran, H., and Shapiro, J. Executive compensation and risk taking. *Review of Finance*, 19(6):2139–2181, 2015.

Bonaccorsi di Patti, E. and Sette, E. Bank balance sheets and the transmission of financial shocks to borrowers: evidence from the 2007-2008 crisis. *Temi di discussione - Bank of Italy Working papers*, Number 848, 2012.

Boot, A. and Thakor, A. Financial system architecture. *Review of Financial Studies*, 10(3): 693–733, 1997.

Boot, A. W. and Thakor, A. V. The Accelerating Interation of Banks and Markets and its Implications for Regulation. In Berger, A. N., Wilson, J. O. S., and Molyneux, P., editors, *The Oxford Handbook of Banking*, chapter 3, pages 59–89. Oxford University Press, Oxford, 2009.

Boot, A. W. A. and Ratnovski, L. Banking and Trading. *Review of Finance*, pages 2219–2246, 2016.

Bos, J. W. B., Economidou, C., and Koetter, M. Technology clubs, R&D and growth patterns: Evidence from EU manufactoring. *European Economic Review*, 54:60–79, 2010a.

Bos, J. W. B., Economidou, C., Koetter, M., and Kolari, J. W. Do all countries grow alike? *Journal of Development Economics*, 91:113–127, 2010b.

Bossone, B. Banks and capital markets: A two-way nexus. *Vox.org*, 2010.

Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., and Sauri, O. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, 196(2):347–367, 2017.

Boyd, J. H., Nicoló, G., and Jalal, A. M. Bank risk-taking and competition revisited: New theory and new evidence. *IMF Working Paper 06/297*, 2006.

Branch, B. The costs of bankruptcy. A review. *International Review of Financial Analysis*, 1: 1–19, 2001.

Bushman, R. M. Transparency, Accounting Discretion, and Bank Stability. *FRBNY Economic Policy Review*, August:129–149, 2016.

Buss, A. and Vilkov, G. Measuring Equity Risk with Option-implied Correlations. *The Review of Financial Studies*, 25(10):3113–3140, 2012.

Campbell, J. Y. Stock returns and the term structure. *Journal of Financial Economics*, 18(2): 373–399, 1987.

Campbell, J. Y. and Hentschel, L. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318, 1992.

Chang, H. S., Donohoe, M., and Sougiannis, T. Do analysts understand the economic and reporting complexities of derivatives? *Journal of Accounting and Economics*, 61:584–604, 2016.

Chranes, J. M., Koch, P., and Berkman, H. Measuring Hedge Effectiveness for FAS 133 Complience. *Journal of Applied Corporate Finance*, 15(4):8–16, 2003.

Cingano, F., Manaresi, F., and Sette, E. Does Credit Crunch Investment Down? New Evidence on the Real Effects of the Bank-Lending Channel. *Review of Financial Studies*, 29(10):2737–2773, 2016.

Cochrane, J. H. Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations. *Journal of Finance*, 46(1):209–237, 1991.

Connolly, R. A. and Rendleman, R. J. J. Properties of Portfolio-Based Estimates of Market Risk Premia. *Working Paper*, 2009.

Cosemans, M., Frehen, R., Schotman, P. C., and Bauer, R. Estimating Security Betas Using Prior Information Based on Firm Fundamentals. *The Review of Financial Studies*, 29(4): 1072–1112, 2016.

Cyree, K. B., Huang, P., and Lindley, J. T. The Economic Consequences of Banks' Derivatives Use in Good Times and Bad Times. *Journal of Financial Services Research*, 41: 121–144, 2012.

Dadalt, P., Gay, G. D., and Nam, J. Asymmetric information and corporate derivatives use. *Journal of Futures Markets*, 22(3):241–267, 2002.

D'Arista, J. W. *The Evolution of U.S. Finance (Volume II)*. M.E. Sharpe, Inc., Armonk, NY, 1994.

Dayton, C. M. and Macready, G. B. Concomitant-Variable Latent-Class Models. *Journal of the American Statistical Association*, 83(401):173–178, 1988.

De Haas, R. and Van Horen, N. Running for the Exit? International Bank Lending During a Financial Crisis. *Review of Financial Studies*, 26(1):244–285, 2012.

Diamond, D. W. Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt. *Journal of Political Economy*, 99(4):689–721, 1991.

Diamond, D. W. and Rajan, R. G. Fear of fire sales, illiquidity seeking, and credit freezes. *The Quarterly Journal of Economcis*, 126(2):557–591, 2011.

Dimson, E. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, (7):197–226, 1979.

Do, C. B. and Batzoglou, S. What is the expectation maximization algorithm? *Nature Biotechnology*, 26:897–899, 2008.

Donohoe, M. Financial derivatives in corporate tax avoidance: a conceptual perspective. *Journal of the American Taxation Association*, 37:37–68, 2015.

Dreher, A. Does globalization affect growth? Evidence from a new index of globalization. *Applied Economics*, 38(10):1091–1110, 2006.

Duffie, D. Market Making Under the Proposed Volcker Rule. *Rock Center for Corporate Governance Stanford University*, Working Paper Serires No. 106, 2012.

Duffie, D. Financial Regulatroy Reform After the Crisis: An Assessment. *Management Science*, pages 1–23, 2017.

Elton, E., Gruber, M., Brown, S., and Goetzmann, W. *Modern Portfolio Theory and Investment Analysis*. Wiley & Sons., Inc., Hoboken, 2003.

Engle, R. F. Dynamic Conditional Correlation. *Journal of Business & Economic Statistics*, 20 (3):339–350, 2002.

Epps, T. W. Comovements in Stock Prices in the Very Short Run. *Journal of the American Statistical Association*, 74(366a):291–298, 1979.

Erel, I., Julio, B., Kim, W., and Weisbach, M. S. Macroeconomic Conditions and Capital Raising. *Review of Financial Studies*, 25(2):689–721, 2011.

Fabozzi, F. J. and Francis, J. C. Beta as a random Coefficient. *Journal of Financial and Quantitative Analysis*, 13(1):101–116, 1978.

Fagerland, M. W. t-tests, non-parametric tests, and large studies—a paradox of statistical practice? *BMC Medical Research Methodology*, 12(78), 2012. doi: http://doi:10.1186/1471-2288-12-78.

Fama, E. and French, K. The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427–465, 1992.

Fama, E. and French, K. R. The Capital Asset Pricing Model: Theory and Evidence. *CRSP Working Paper No. 550; Tuck Business School Working Paper No. 03-26*, 2003.

Fama, E. F. and MacBeth, J. D. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3):607–636, 1973.

Fay, M. P. and Proschan, M. P. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, 4:1–39, 2010. doi: http://doi.org/10.1214/09-SS051.

Federal Reserve Bank of New York. Crsp-frb link, 2017. URL `https://www.newyorkfed.org/research/banking_research/datasets.html`.

FinancialCAD & KPMG. *Basics of Hedge Effectiveness Testing and Measurement*. FinancialCAD Corporation, Surrey, Canada, 2011.

Frazzini, A. and Pedersen, L. H. Betting against beta. *Journal of Financial Economics*, 111: 1–25, 2014.

French, K. R., Schwert, G. W., and Stambaugh, R. F. Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29, 1987.

Froot, K. A., Scharfstein, D. S., and Stein, J. C. Risk Management: Coordinating Corporate Investment and Financing Policies. *Journal of Finance*, 48(5):1629–1658, 1993.

Gebhardt, G., Reichenhardt, R., and Wittenbrink, C. Accounting for financial instruments in the banking industry: conclusions from a simulation model. *European Accounting Review*, 13(2):341–371, 2004.

Gensler, G. OTC Derivatives Reform. In *Keynote Address of Chairman Gary Gensler at the Markit's Outlook for OTC Derivatives Markets Conference*. New York, N.Y., March 9, 2010.

Ghysels, E., Santa-Clara, P., and Valkanov, R. There is a risk-return trade-off after all. *Journal of Financial Economics*, 76:509–548, 2005.

Glosten, L. R. and Milgrom, P. R. Bid, ask and transaction prices in a specialist market with heterogeneously infromed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48 (5):1779–1801, 1993.

Gobat, J. What Is a Bank? *Finance & Development*, 49(1):38–39, 2012.

Goldstein, M. and Xie, D. The Impact of the Financial Crisis on Emerging Asia. *Peterson Institute for International Economics Working Paper Series*, WP 09-11, 2009.

Graham, J. R. and Harvey, C. R. The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics*, 60:187–243, 2001.

Grauer, R. R. and Janmaat, J. A. On the power of cross-sectional and multivariate tests of the CAPM. *Journal of Banking & Finance*, 33:775–787, 2009.

Greene, W. H. *Econometric Modeling Guide Vol. 1*. Econometric Sorftware, Inc., New York, 2002.

Greenwald, B. and Stiglitz, J. Financial market imperfections and business cycles. *Quarterly Journal of Economics*, 108:77–104, 1993.

Grün, B. and Leisch, F. FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. *Journal of Statistical Software*, 28(4), 2008. doi: 10.18637/jss.v028.i04.

Gygli, S., Haelg, F., and Sturm, J.-E. The KOF Globalisation Index - Revisited. *KOF Working Papers No. 439*, 2018.

Hardy, D. C. Bank Resolution Costs, Depositor Preference, and Asset Encumberance. *IMF Working Paper WP/13/172*, 2013.

Harvey, C. R. The specification of conditional expectations. *Journal of Empirical Finance*, 8 (5):573–637, 2001.

Hasbrouck, J. and Ho, T. S. Y. Order Arrival, Quote Behavior, and the Return-Generating Process. *Journal of Finance*, 42(4):1035–1048, 1987.

Haugen, R. A. and Heins, A. J. Risk and the Rate of Return on Financial Assets: Some Old Wine in New Bottles. *Journal of Financial and Quantitaive Analysis*, 10(5):775–784, 1975.

Hayashi, T. and Yoshida, N. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.

Hu, J. Does option trading convey stock price information? *Journal of Financial Economics*, (11):625–645, 2014.

Hughen, L. When Do Accounting Earnings Matter More than Econommic Earnings? Evidence from Hedge Accounting Restatements. *Journal of Business Finance & Accounting*, 37(9):1027–1056, 2010.

Ivashina, V. Structure and Pricing of Syndicated Loans. In *The New York City Area Conference on Financial Intermediation jointly sponsored by the Federal Reserve Bank of New York and the Salomon Center*. New York, N.Y., November 18, 2005.

Ivashina, V. and Scharfstein, D. Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, 97:319–338, 2010.

Iyer, R., Peydro, J.-L., da Rocha-Lopes, and Schoar, A. Interbank Liquidity Crunch and the Firm Credit Crunch: Evidence from the 2007-2009 Crisis. *Review of Financial Studies*, 27 (1):347–372, 2014.

Jagannathan, R. and Wang, Z. The Conditional CAPM and the Cros-Section of Expected Returns. *Journal of Finance*, 51(1):3–53, 1996.

Jarmin, R. S. Learning by Doing and Competition in the Early Rayon Industry. *The RAND Journal of Economics*, 25(3):441–454, 1994.

Jimenez, G., Ongena, S., Peydro, J.-L., and Saurina, J. Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications. *American Economic Review*, 102(5):2301–2326, 2012.

Jimenez, G., Ongena, S., Peydro, J.-L., and Saurina, J. Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking? *Econometrica*, 82(2):463–505, 2014.

Jorion, P. *Financial Risk Manager Handbook*. John Wiley & Sons Inc., Hoboken, New Jersey, 2nd edition, 2003.

Kahane, Y. Capital adequacy and the regulation of financial intermediaries. *Journal of Banking & Finance*, 1:207–218, 1977.

Kahle, K. M. and Stulz, R. M. Access to Capital, Investment, and the Financial Crisis. *Journal of Financial Economics*, 110(2):280–299, 2013.

Karlis, D. and Xekalaki, E. Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis*, 41:577–590, 2003.

Kawaller, I. What analyst need to know about accounting for derivatives. *Financial Analyst Journal*, 60:24–30, 2004.

Keat, H. S. The Global Financial Crisis: Impact on Asia and Poicy Challenges Ahead. *Federal Reserve Bank of San Francisco Proceedings*, pages 267–276, 2009.

Keeley, M. C. Deposit Insurance, Risk, and Market Power in Banking. *American Economic Review*, 80(5):1183–1200, 1990.

Khwaja, A. I. and Mian, A. Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market. *American Economic Review*, 98:1413–1442, 2008.

Kim, D. The Errors in the Variables Problem in the Cross-Section of Expected Stock Returns. *Journal of Finance*, 50(5):1605–1634, 1995.

Köhler, M. Which banks are more risky? The impact of business models on bank stability. *Journal of Financial Stability*, 16:195–212, 2015.

Krahnen, J.-P., Noth, F., and Schüwer, U. Structural Reforms in Banking: The Role of Trading. *Journal of Financial Regulation*, 3:66–88, 2017.

Kyle, A. S. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, 1985.

Laeven, L. and Levine, R. Bank governance, regulation and risk taking. *Journal of Financial Economics*, 93:259–275, 2009.

Lehmann, M. Volcker rule, ring-fencing or separation of bank activities - Comparison of structural reform acts around the world. *Journal of Banking Regulation*, 17(3):176–187, 2016.

Leone, M. Hedge Accounting: A Matchless Future? *CFO.com*, May 24, 2007.

Lepetit, L. and Strobel, F. Bank insolvency risk and Z-score measures: A refinement. *Finance Research Letters*, 13:214–224, 2015.

Lesmond, D. A. Liquidity of emerging markets. *Journal of Financial Economics*, 77(2): 411–452, 2005.

Li, X., Sullivan, R. N., and Garcia-Feijoo. The Limits to Arbitrage and the Low-Volatility Anomaly. *Financial Analysts Journal*, 70(1):52–64, 2014.

Lintner, J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.

Lo, A. W. and MacKinlay, A. C. Data-Snooping Biases in Tests of Financial Asset Pricing Models. *Review of Financial Studies*, 3(3):431–467, 1990.

Magidson, J. and Vermunt, J. Latent class models. In Kaplan, D., editor, *The Sage handbook of quantitative methodology for the social sciences*, pages 175–198. Sage, Thousands Oakes, 2004.

Marchetti, J. The International Banking Landscape: Developments, Drivers, and Potential Implications. In Demirgüc-Kunt, A., Evanoff, D. D., and Kaufman, G. G., editors, *The Future of Large, Internationally Active Banks*, chapter 7, pages 97–111. World Scientific Studies in International Economics 55, 2016.

Marcus, A. J. Deregulation and Bank Financial Policy. *Journal of Banking & Finance*, 8: 557–565, 1984.

Mayers, D. and Smith, C. W. On the Corporate Demand for Insurance. In Dionne, G. and Harrington, S. E., editors, *Foundations of Insurance Economics*, volume 14 of *Huebner International Series on Risk*. Springer, Dordrecht, 1982.

Mayordomo, S., Rodriguez-Moreno, M., and Pena, J. I. Derivatives holdings and systemic risk in the U.S. banking sector. *Journal of Banking and Finance*, 45:84–104, 2014.

Mercieca, S., Schaeck, K., and Wolfe, S. Small european banks: Benefits from diversification? *Journal of Banking & Finance*, 31(7):1975–1998, 2007.

Mergaerts, F. and Vennet, R. V. Business models and bank performance: A long-term perspective. *Journal of Financial Stability*, 22:57–75, 2016.

Merkley, J. and Levin, C. The Dodd-Frank Act restrictions on proprietary trading and conflicts of interest: new tools to address evolving threats. *Harvard Journal on Legislation*, 48:515–553, 2011.

Merton, R. C. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2), 1974.

Merton, R. C. An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees. *Journal of Banking & Finance*, 1:3–11, 1977.

Merton, R. C. Applications of Option-Pricing Theory: Twenty-Five Years Later. *Nobel Lecture*, 1997.

Miles, D., Yang, J., and Marcheggiano, G. Optimal Bank Capital. *The Economic Journal*, 123: 1–37, 2012.

Modigliani, F. and Miller, M. The cost of capital, corporate finance and the theory of investment. *American Economic Review*, 48:261–297, 1958.

Monin, P. The OFR Financial Stress Index. *Office of Financial Research Working Paper*, 17-04, 2017.

Morey, M. R. Have Leading Finance Textbooks Incorporated Recent Events? *Journal of Financial Education*, 41(3):26–46, 2015.

Morrison, A. D. Universal Banking. In Berger, A. N., Wilson, J. O. S., and Molyneux, P., editors, *The Oxford Handbook of Banking*, chapter 7, pages 171–194. Oxford University Press, Oxford, 2009.

Mulford, C. W. and Comiskey, E. E. The Non-designation of Derivatives as Hedges for Accounting Purposes. *Journal of Applied Research in Accounting and Finance*, 3(2):3–15, 2008.

Naes, R., Skjeltorp, J. A., and Odegaard, A. Stock Market Liquidity and the Business Cycle. *Journal of Finance*, 66(1):139–176, 2011.

Nelson, D. B. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347–370, 1991.

Novy-Marx, R. and Velikov, M. A Taxonomy of Anomalies and Their Trading Costs. *Review of Financial Studies*, 29(1):104–147, 2016.

Papa, V. T. and Peters, S. J. Derivatives and Hedging Activities Disclosures. In *User Perspectives on Financial Instrument Risk Disclosures under International Financial Reporting Standards*, volume 2. CFA Institute, 2013.

Paravisini, D., Rappoport, V., and Schnabl, P. Specialization in Bank Lending: Evidence from Exporting Firms. *NBER Working Paper No 21800*, 2015.

Park, S. Risk-taking behavior of banks under regulation. *Journal of Banking & Finance*, 21: 491–507, 1997.

Patton, A. J. and Weller, B. M. Risk Price Variation: The Missing Half of the Cross-Section of Expected Returns. *ERID Working Paper Number 274*, 2018.

Pennings, J. M. E. and Garcia, P. Hedging behavior in small and medium-size enterprises: The role of unobserved heterogeneity. *Journal of Banking & Finance*, 28:951–978, 2004.

Pesaran, M. H. and Timmermann, A. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161, 2007.

Purnanandam, A. Interest rate derivatives at commercial banks: An empirical investigation. *Journal of Monetary Economics*, 54:1769–1808, 2007.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., and William, T. R. An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data. *Marketing Science*, 12(1):103–124, 1993.

Rampini, A. A., Viswanathan, S., and Vuillemey, G. Risk Management in Financial Institutions. *Working Paper*, 2017.

Reno, R. A Closter Look at the Epps Effect. *International Journal of Theoretical and Applied Finance*, 6(1):87–102, 2003.

Reuters Business News. U.S. considering 'material changes' to 'Volcker Rule': Fed's Quarles. `https://www.reuters.com/article/us-usa-fed-quarles/u-s-considering-material-changes-to-volcker-rule-feds-quarles-idUSKBN1GH2U8`, 2018. Accessed: 2018-05-21.

Rosenberg, B. and Guy, J. Beta as a random Coefficient. *Financial Analyst Journal*, May-June:60–72, 1976.

Scholes, M. and Williams, J. Estimating betas from nonsynchronous data. *Journal of Financial Economics*, (5):309–327, 1977.

Schwert, M. Bank Capital and Lending Relationships. *Journal of Finance*, 73(2):787–830, 2018.

Shanken, J. and Zhou, G. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics*, 84:40–86, 2007.

Shapiro, A. and Titmann, S. An integrated approach to corporate risk management. In Stern, J. M. and Chew Jr, D. H., editors, *The Revolution in Corporate Finance*, pages 331–354. Blackwell Business, Malden, 1998.

Sharpe, W. F. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*, 19(3):425–442, 1964.

Shleifer, A. and Vishny. Unstable banking. *Journal of Financial Economics*, 97:306–318, 2010.

Shumway, S. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74:101–124, 2001.

Simpson, E. H. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society*, Series B(13):238–241, 1951.

Sinkey, J. F. J. and Carter, D. A. Evidence on the financial characteristics of banks that do and do not use derivatives. *The Quarterly Review of Economics and Finance*, 40:431–449, 2000.

Smith, C. W. and Stulz, R. M. The Determinads of Firms' Hedging Policies. *Journal of Financial and Quantitative Analysis*, 20(4):391–405, 1985.

Song, F. and Thakor, A. Financial system architecture and the co-evolution of banks and capital markets. *Economic Journal*, 120(547):1021–1055, 2010.

Stein, J. C. The Fire-Sales Problem and Securities Financing Transactions. In *Speech by Governor Jeremy C. Stein at the Federal Reserve Bank of New York Workshop on Fire Sales as a Driver of Systemic Risk in Triparty Repo and other Secured Funding Markets*. New York, N.Y., October 4, 2013.

Sufi, A. Information asymetrie and financing arragements: Evidence from syndicated loans. *Journal of Finance*, 62:629–668, 2007.

Thompson, P. Learning by Doing. In Hall, B. H. and Rosenberg, N., editors, *Handbook of The Economics of Innovation*, volume 1, chapter 10, pages 429–476. Elsevier, 2010.

Turner, C. M., Startz, R., and Nelson, C. R. A Markov Model of Heteroskedasticity, Risk, and Learing in the Stock Market. *Journal of Financial Economics*, 25:3–22, 1989.

Valladares, M. Derivatives Transparency Is Essential to End TBTF. *The New York Times*, May 13, 2014.

Vasicek, O. A note on using cross-sectional information in Baysian estimation on security betas. *Journal of Finance*, 28(5):1233–1239, 1973.

Veredas, D. and Petkovic, A. Aggregation in linear models for panel data. *Journal of the Japan Statistical Society*, 40(1):63–95, 2010.

Wedel, M. and DeSarbo, W. A Mixture Likelihood Approach for Generalized Linear Models. *Journal of Classification*, pages 21–55, 1995.

Welch, I. *Corporate Finance. An Introduction*. Pearson Education, Boston, 2009.

Whitehead, C. K. The Volcker Rule and Evolving Financial Markets. *Harvard Business Law Review*, 1:39–73, 2011.

Worstall, T. Banning Prop Trading By Banks: At Some Point It Becomes Banning Banking. `https://www.forbes.com/sites/timworstall/2013/03/24/banning-prop-trading-by-banks-at-some-point-it-becomes-banning-banking/#967ff4571f8a`, 2013. Accessed: 2018-05-21.

Zhang, L. The Value Premium. *Journal of Finance*, IX(1):67–103, 2005.

Curriculum Vitae

Michael Kurz was born on November 20, 1988 in Esslingen am Neckar. He earned his BSc in Business Administration and Economics from Hohenheim University. After completing his BSc, Michael completed two internships, one at the Privatbank Ellwanger & Geiger KG and another at the EADS subsidiary Tesat Spacecom GmbH & Co.KG. He then continued to pursue his MSc in Economics and Finance with a PhD-orientation from Eberhard Karls University of Tübingen. As master's degree student he worked as teaching assistant for the Chair of Statistics and Econometrics at Eberhard Karls University of Tübingen School of Business and Economics. Michael also completed the Maastricht Summer School in Country Risk Analysis and the Big Data in Finance and International Macroeconomics Summer School at the Kiel Institute for the World Economy.

In October 2015, Michael joined the Department of Finance in the School of Business and Economics at Maastricht University to pursue his PhD. As a PhD candidate he taught courses in Options & Futures, Corporate Finance, and Institutional Investors at the BSc- and MSc-level and supervised multiple bachelor's and master's degree theses.

Michael's research has been presented at international conferences and seminars, including the Siena Finance Workshop at Siena University, the annual meeting of the German Finance Association (Deutsche Gesellschaft für Finanzwirtschaft) at Ulm University, the International PhD Colloquium at the University of Luxembourg, Dongbei University of Finance and Economics in Dalian (China), the annual conference of the Scottish Economic Society in Perth (UK), the summer meeting of the International Banking, Economics, and Finance Association (IBEFA) in Vancouver, and the Corporate Finance Day at Antwerp University.

In January 2019, Michael joined the Financial Markets Division of De Nederlandsche Bank (central bank of the Netherlands).